

26 March, 2017 updated V2

# CONCEPTUAL DESIGN REPORT FOR THE US HL-LHC UPGRADE OF THE COMPACT MUON SOLENOID

DRAFT

DRAFT



## Contents

<b>1</b>	<b>Executive Summary - Editor O'Dell</b>	<b>2 pages</b>	<b>1</b>
<b>2</b>	<b>Project Overview - Editor O'Dell, Hill, Snow, Spalding</b>	<b>30 pages</b>	<b>3</b>
2.1	Overview of the HL-LHC Physics Program		3
2.2	Overview of the CMS Experiment and the Upgrades needed for HL-LHC		6
2.2.1	Radiation Damage to the CMS Detector at the HL-LHC		7
2.2.2	High Pileup		9
2.2.3	A brief introduction to the CMS detector		11
2.2.4	General Considerations for the Phase-II upgrade		12
2.2.5	Elements of the Phase-II Upgrades		15
2.2.6	Upgrade Performance Studies		17
2.3	Project Organization and Management		18
2.4	Key Performance Parameters		18
2.5	Cost and Schedule		18
<b>3</b>	<b>Science - Editor Hill</b>	<b>30 pages</b>	<b>19</b>
3.1	Physics Goals to Physics Requirements		19
3.2	Precision Measurements of the Higgs Couplings		19
3.2.1	$H \rightarrow ZZ^* \rightarrow 4\ell$ analysis		20
3.2.2	$H \rightarrow \mu\mu$ analysis		20
3.2.3	$H \rightarrow \tau\tau$ analysis		21
3.2.4	Higgs Cross Section Measurements		21
3.2.5	Higgs boson pair production		22
3.3	Searches for SUSY BSM, including DM		25
3.3.1	Overview of SUSY search strategies and full-spectrum models		28
3.3.2	Results from sensitivity studies		30
3.3.3	SUSY conclusions		35
3.4	Non-SUSY searches for DM		36
3.5	Studying Rare SM Processes		37
3.5.1	Same-sign WW scattering		39
3.5.2	WZ scattering		40
3.5.3	B physics		40
3.6	Physics Requirements		41
3.6.1	Durable detector capable of high performance until $3 \text{ ab}^{-1}$ integrated		42
3.6.2	Efficient reconstruction for all Higgs Decays		42
3.6.3	Ability to disentangle different Higgs Production		42
3.6.4	Electroweak scale single-lepton trigger thresholds		42
3.6.5	Percent level lepton momentum resolution		42
3.6.6	Efficient b-tagging in central and forward Rapidities		42
3.6.7	Efficient VBF jet tagging and associated primary vertex resolution		42
3.6.8	Excellent missing transverse energy resolution		42
3.6.9	Pile-up mitigation such that LHC performance recovered at HL-LHC		42

50	<b>4 Outer Tracker - Editor Canepa 40 pages</b>	<b>45</b>
51	4.1 Limitations of the Phase 1 tracking system . . . . .	45
52	4.2 Requirements for the Phase 2 tracker system . . . . .	46
53	4.3 Brief overview of the Phase 2 tracker system . . . . .	48
54	4.4 Overview of the Outer Tracker . . . . .	50
55	4.4.1 Outer tracker sensors . . . . .	50
56	4.4.2 Outer tracker modules . . . . .	54
57	4.4.3 Outer tracker electronics . . . . .	58
58	4.4.4 Outer tracker mechanical structure . . . . .	64
59	4.4.5 Outer tracker prototyping, system tests, and beam tests . . . . .	68
60	4.4.6 Outer tracker expected physics performance . . . . .	72
61	4.5 US CMS contributions . . . . .	81
62	4.5.1 Sensors . . . . .	81
63	4.5.2 Modules . . . . .	81
64	4.5.3 Electronics . . . . .	93
65	4.5.4 Flat Barrel (Flat TBPS) . . . . .	96
66	4.5.5 Prototyping, system tests, and beam tests . . . . .	101
67	<b>5 Barrel Hadronic Calorimeter - Editor Jessop 15 Pages</b>	<b>105</b>
68	5.1 Introduction . . . . .	105
69	5.2 Requirements . . . . .	105
70	5.3 Proposed design . . . . .	105
71	5.3.1 Scintillator choice . . . . .	106
72	5.3.2 Optical Path . . . . .	107
73	5.3.3 Back-end Electronics . . . . .	107
74	5.4 Performance . . . . .	107
75	5.5 Alternatives . . . . .	107
76	5.5.1 Scintillator and WLS alternatives . . . . .	108
77	5.5.2 Segmentation . . . . .	108
78	<b>6 Endcap Calorimeter - Editors Yohay Mans 40 pages</b>	<b>111</b>
79	6.1 Introduction . . . . .	111
80	6.2 Requirements . . . . .	114
81	6.3 Proposed design . . . . .	118
82	6.3.1 Design overview . . . . .	118
83	6.3.2 Cassettes . . . . .	119
84	6.3.3 Silicon modules . . . . .	123
85	6.3.4 Scintillator modules . . . . .	126
86	6.3.5 Mechanical design . . . . .	126
87	6.3.6 Electronics and electrical systems . . . . .	126
88	6.3.7 Calibration and monitoring . . . . .	134
89	6.4 US deliverables . . . . .	135
90	6.4.1 Silicon sensors and modules . . . . .	135
91	6.4.2 Scintillator modules . . . . .	135

92	6.4.3	Hadron calorimeter cassettes . . . . .	135
93	6.4.4	Electronics and electrical systems . . . . .	135
94	6.5	Performance . . . . .	135
95	6.5.1	Design energy resolution . . . . .	136
96	6.5.2	MIP sensitivity . . . . .	136
97	6.5.3	Occupancy . . . . .	137
98	6.5.4	Test beam results . . . . .	137
99	6.6	Alternatives . . . . .	137
100	6.7	R&D Towards Final Design . . . . .	137
101	6.7.1	Radiation hard plastic scintillator . . . . .	137
102	6.7.2	Trigger data compression . . . . .	139
103	6.7.3	Cassette prototyping program . . . . .	139
104	6.8	Risk . . . . .	139
105	6.9	Value Engineering . . . . .	139
106	6.10	Quality Assurance and Quality Control . . . . .	139
107	6.10.1	Sensor and Module Quality Assurance and Quality Control . . . . .	139
108	6.11	Environmental Impact, Health, and Safety . . . . .	141
109	<b>7</b>	<b>Trigger and DAQ - Editor Berryhill 40 pages</b>	<b>143</b>
110	7.1	Introduction . . . . .	143
111	7.2	Requirements . . . . .	144
112	7.3	Proposed design . . . . .	146
113	7.3.1	Architecture Overview . . . . .	146
114	7.3.2	Calorimeter trigger . . . . .	146
115	7.3.3	Muon trigger . . . . .	148
116	7.3.4	Track trigger . . . . .	150
117	7.3.5	Correlator trigger . . . . .	151
118	7.3.6	DAQ . . . . .	151
119	7.4	Performance . . . . .	153
120	7.4.1	Calorimeter trigger with crystal granularity . . . . .	153
121	7.5	Alternatives . . . . .	154
122	7.6	Environmental Impact, Health, and Safety . . . . .	155
123	7.7	Risk . . . . .	156
124	7.8	Quality Assurance and Quality Control . . . . .	156
125	7.9	Value Engineering . . . . .	157
126	7.10	R&D Towards Final Design . . . . .	157
127	<b>8</b>	<b>CSC Muon Upgrade (DOE scope) - Editor Stan Durkin</b>	<b>159</b>
128	8.1	Introduction . . . . .	159
129	8.2	Requirements . . . . .	161
130	8.3	Proposed design . . . . .	162
131	8.3.1	CFEB Rate Problems . . . . .	162
132	8.3.2	DCFEBBoards . . . . .	163
133	8.3.3	OTMB Boards . . . . .	165

134	8.3.4	ALCT Mezzanine Boards . . . . .	166
135	8.3.5	Links . . . . .	166
136	8.3.6	LVDB . . . . .	166
137	8.3.7	Staging . . . . .	166
138	8.3.8	Infrastructure . . . . .	167
139	8.4	Performance . . . . .	167
140	8.5	Alternatives . . . . .	168
141	8.6	Environmental Impact, Health, and Safety . . . . .	168
142	8.7	Risk . . . . .	169
143	8.8	Quality Assurance and Quality Control . . . . .	169
144	8.9	Value Engineering . . . . .	170
145	8.10	R&D Towards Final Design . . . . .	170
146	<b>9</b>	<b>CSC Muon Upgrade (non DOE scope) - Editor Alexei Safonov</b>	<b>171</b>
147	9.1	Introduction . . . . .	171
148	9.2	Requirements . . . . .	172
149	9.3	Proposed design . . . . .	173
150	9.4	Performance . . . . .	175
151	9.5	Alternatives . . . . .	175
152	9.6	Environmental Impact, Health, and Safety . . . . .	176
153	9.7	Risk . . . . .	176
154	9.8	Quality Assurance and Quality Control . . . . .	176
155	9.9	Value Engineering . . . . .	177
156	9.10	R&D Towards Final Design . . . . .	177
157	<b>10</b>	<b>Summary - Editor (TBD) 2 pages</b>	<b>179</b>
158	.1	Appendix: Outer tracker common mechanics and services . . . . .	179
159	.1.1	Mechanics . . . . .	179
160	.1.2	Services . . . . .	180
161	.1.3	Tracker assembly sequence . . . . .	181
162	.1.4	Cooling system . . . . .	182
163	.1.5	Central beampipe . . . . .	183
164	<b>A</b>	<b>Endcap Calorimeter Supplementary Material</b>	<b>185</b>
165	A.1	Cassette layout information . . . . .	185
166		<b>References</b>	<b>193</b>

# Figures

2.1	Observed and projected precision on Higgs boson couplings as a function of boson or fermion masses. . . . .	4
2.2	SUSY mass limits from the Run I dataset . . . . .	5
2.3	Projected LHC performance through 2035, showing preliminary dates for long shutdowns of LHC and projected luminosities. . . . .	6
2.4	Pseudorapidity and transverse momentum distributions of the particles generated by the DPMJET III generator. The black line is for all particles, the blue line is for charged particles. . . . .	8
2.5	Absorbed dose in the CMS cavern after an integrated luminosity of $3000 \text{ fb}^{-1}$ . R is the transverse distance from the beamline and Z is the distance along the beamline from the Interaction Point at $Z=0$ . . . . .	9
2.6	High pileup event with 78 reconstructed vertices taken in 2012 . . . . .	11
2.7	Illustration of out-of-time pileup: (left) Individual pulses from a detector as a function of time in ns. The “triggered” or “signal” BX is located at “0” and is bounded by the green vertical lines. The “signal” pulse appears in blue and extends into the next two 25ns “buckets”, which contain some pulse height from the later pileup collisions. Some energy from the preceding “bucket” also falls into the “signal” BX. (right) The sum of the pulse heights as a function of time. The challenge is to correct the “summed” pulse height or charge (related to area under the pulse height distribution) to derive the pulse height or the total charge in the blue signal pulse. . . . .	11
2.8	A schematic representation of the CMS Detector in Run 1, with its various sections in retracted positions. The central yoke block is YB0. The first block (YB+1, and corresponding YB-1) is shown partially moved away from YB0. The second yoke block (YB+2, which has a corresponding YB-2) is shown fully moved past the solenoid vacuum tank. The endcap calorimeters are attached to the first endcap disk YE+1, then endcap CSCs and RPCs, then YE+2, more muon chambers, and YE+3, with additional muon chambers on the front and back. Eventually, another disk, YE+4 will be added at the end to provide shielding from beam related backgrounds. This configuration is repeated on the other end. In operation, the detector is closed by moving all pieces together. . . . .	13

200	3.1	Estimated precision on the measurements for modified couplings for a SM-like Higgs boson [4]. The projections assume a center-of-mass energy of 14 TeV and a dataset with integrated luminosity of $3000 \text{ fb}^{-1}$ . The projections are obtained with two uncertainty scenarios as described in the text. . . . .	20
201			
202			
203			
204	3.2	The projected relative uncertainties for the $H \rightarrow \gamma\gamma$ fiducial cross section ( $\sigma_{\text{fid}}$ ) are shown. For the $300 \text{ fb}^{-1}$ case, the scenarios are S1 and S2, while for the $3000 \text{ fb}^{-1}$ case, the scenarios are S1+ and S2+. In each case the uncertainty for the inclusive signal strength is also shown split into components: statistical uncertainties ('stat.'), experimental systematic uncertainties ('exp.'). . . . .	23
205			
206			
207			
208			
209	3.3	Projections for the differential cross section measurement of the Higgs boson transverse momentum at $300 \text{ fb}^{-1}$ () and $3000 \text{ fb}^{-1}$ (). The last bin represents the integrated cross section for $p_T(H) > 200 \text{ GeV}$ and is normalized by 1/50 for presentation. The theoretical uncertainty on the differential gluon fusion cross section is shown in magenta. . . . .	23
210			
211			
212			
213			
214	3.4	Feynman diagrams contributing to gluon fusion Higgs boson pair production.	24
215	3.5	Di-photon mass distribution for the estimated signal and background contributions. The data points show the result of a pseudo-experiment. . . . .	24
216			
217	3.6	Examples of SUSY full-spectrum models: (a) the natural SUSY model NM3 and (b) the stau coannihilation model STC, which are among the five full-spectrum scenarios used in the studies presented here. In NM3, the masses of the $\tilde{g}$ , $\tilde{t}_1$ , $\tilde{t}_2$ , and $\tilde{b}_1$ are all below 2 TeV. The $\tilde{\chi}_1^0$ is higgsino-like. In the STC model, the gluino is much heavier than the top squarks, and the slepton sector is light, with the $\tilde{\tau}$ nearly degenerate with the $\tilde{\chi}_1^0$ . The lines between different states indicate transitions with branching fractions greater than 5%. . . . .	27
218			
219			
220			
221			
222			
223			
224	3.7	Simplified SUSY models for chargino-neutralino production leading to (a) $W^\pm Z + E_T^{\text{miss}}$ and (b) $W^\pm H + E_T^{\text{miss}}$ final states. . . . .	27
225			
226	3.8	Distributions from two searches with sensitivity to the NM1, NM2, and NM3 models. (a) All-hadronic search with $M_{T2}$ : the distribution of $M_{T2}$ after the full selection (including $\geq 3$ b-tags) except the $M_{T2}$ requirement itself. (b) Single-lepton search: distribution of $H_T$ after the selection requirements are applied to all other variables. The contributions of the SM backgrounds are shown as stacked histograms in both (a) and (b), as they are elsewhere in this paper. In (a), the SUSY signal contributions from the different models are shown overlaid, but in (b) they are shown stacked because the histograms represent different processes within a single SUSY scenario, NM1. . . . .	31
227			
228			
229			
230			
231			
232			
233			
234			
235	3.9	Trilepton searches with and without b-jet tagging. (a) Distribution of $E_T^{\text{miss}}$ for the trileptons + b-jets search in the bin with $\geq 4$ b-tagged jets. (b) Distribution of $E_T^{\text{miss}}$ in the trilepton search for $\tilde{\chi}_1^\pm \tilde{\chi}_1^0$ , with $\tilde{\chi}_1^\pm \rightarrow W^\pm \tilde{\chi}_1^0$ and $\tilde{\chi}_2^0 \rightarrow Z \tilde{\chi}_1^0$ . In this search, no b-tagging requirements are used. The selected events satisfy $200 < M_T < 400 \text{ GeV}$ and have an $\ell^+ \ell^-$ pair that reconstructs to the Z-boson mass. . . . .	32
236			
237			
238			
239			
240			
241	3.10	(a) Distribution of the dilepton invariant mass $m_{\ell^+ \ell^-}$ in NM1, showing the distinctive kinematic edge associated with the decay chain $\tilde{\chi}^\pm \rightarrow \tilde{\ell}^\pm \ell^\mp$ ; $\tilde{\ell}_L^\pm \rightarrow \ell^\pm \tilde{\chi}_1^0$ . (b) Search for $\tilde{\chi}_2^0 \tilde{\chi}_1^\pm \rightarrow WH + E_T^{\text{miss}}$ : distribution of $E_T^{\text{miss}}$ . . . . .	32
242			
243			

244	3.11	Search for $\tilde{\chi}_2^0 \tilde{\chi}_1^\pm$ production in the $W^\pm Z + E_T^{\text{miss}}$ and $W^\pm H + E_T^{\text{miss}}$ final state. The excluded regions are shown in the simplified model parameter space of $m(\tilde{\chi}_1^0)$ vs. $m(\tilde{\chi}_1^\pm) = m(\tilde{\chi}_2^0)$ for various assumptions. In such plots, the mass of the produced particle (or particles) is generally shown on the $x$ -axis, while the mass of the LSP is shown on the $y$ -axis. As a consequence, the excluded region is bounded by the decreasing production cross section on the right, but by the decreasing $E_T^{\text{miss}}$ as one approaches the diagonal. . . . .	34
251	3.12	All-hadronic $\tilde{b}_1$ search. Distributions of the $M_{CT}$ variable for (a) $M_T > 750$ GeV and (b) $M_T > 950$ GeV. The endpoint of the $M_{CT}$ distribution is a function of $m(\tilde{b}_1)$ and $m(\tilde{\chi}_1^0)$ . . . . .	34
254	3.13	Monojet-like search in STOC: distribution of the $p_T$ of the leading jet in the event. . . . .	35
256	3.14	All-hadronic $H_T - H_T^{\text{miss}}$ search. (a) Distribution of $H_T^{\text{miss}}$ for the SM backgrounds and the STC and STOC signal contributions. (b) Distribution of the number of b-tagged jets after all other selection requirements. . . . .	35
259	3.15	Dark matter reach of the monolepton channel as a function of the DM mass and mediator mass for the two extreme cases of $\zeta = \pm 1$ . . . . .	37
261	3.16	Left, the $\Delta\phi$ between the two final state charged leptons for the same-sign WW scattering, after the VBS selections, for positive muons in the non-aged Phase-I scenario. Right, an example of the expected differences for polarized scatterings in the WZ analysis. . . . .	38
265	3.17	Left, the differences in shape of $\Delta\eta_{jj}$ between the two final state jets for the WW scattering, before the VBS selections, for signal and background. Right, the $m_{\ell\ell}$ distribution at the end of the analysis chain, still for the WW scattering, showing the expectation for the SM case and for a signal hypothesis with enhanced anomalous couplings. . . . .	39
270	3.18	Expected 95% CL 2D contour for the $S_0$ and $S_1$ parameters (right). . . . .	40
271	3.19	Projections of the mass fits to $3000 \text{ fb}^{-1}$ of integrated luminosity, respectively assuming the expected performances of upgraded CMS detector. The plot is for the barrel only. . . . .	41
274	4.1	Sketch of one quarter of the Run 1 CMS tracking system. The radial region below 200 mm is equipped with pixellated detectors. Beyond 200 mm, the outer tracker features single-sided strip modules (red segments) and double-sided modules composed by two back-to-back silicon strip detectors with a stereo angle of 100 mrad (blue segments). Double-sided modules provide measurements of the $z$ and $r$ coordinates in the barrel and end-cap, respectively. The four inner tracker barrel layers (six outer tracker barrel layers) are referred to as TIB (TOB). The three inner forward disks (nine end cap disks) on each side are referred to as TID $D_i$ (TEC $D_i$ ). <b>FIXME:</b> This plot may be replaced, waiting for final version to be included in the TDR. . . . .	46



284	4.2	Map of non-functional modules (in blue) after an integrated luminosity of $1000 \text{ fb}^{-1}$ , for the achievable minimum coolant temperature of $-20^\circ\text{C}$ . Almost all the stereo modules in the barrel (Inner Barrel layers 1 and 2 and Outer Barrel layers 1 and 2), as well as in the endcap (rings 1, 2 and 5), are no longer operational. The evolution of the leakage current of the tracker sensors is predicted by a detailed model that takes into account the estimated luminosity profile, the position and size of each module, the expected particle fluence at specific module locations (obtained from FLUKA simulations and the expected temperature versus time scenario that includes annealing periods. The model also implements a map of the efficiency of the module thermal contacts derived from data. . . . .	47
295	4.3	Reconstruction efficiency for $p_T = 10 \text{ GeV}$ muons as a function of pseudorapidity for the Phase-I tracker before and after the tracker has been aged by an equivalent integrated luminosity of $1000 \text{ fb}^{-1}$ . . . . .	47
298	4.4	Integrated particle fluence in $1 \text{ MeV}$ neutron equivalent per $\text{cm}^2$ , for the Phase 2 tracker. The estimates shown here correspond to a total integrated luminosity of $3000 \text{ fb}^{-1}$ of $pp$ collisions at $\sqrt{s} = 14 \text{ TeV}$ . <b>FIXME:</b> Plot needs to be updated. TID plot to be added. Update geometry overlay. . . . .	49
302	4.5	Sketch of one quarter of the Phase 2 tracker layout in $r$ - $z$ view. In the Inner Tracker the green lines correspond to pixel modules made of two readout chips and yellow lines to pixel modules with four readout chips. In the Outer tracker the blue lines correspond to PS modules, while red lines correspond to 2S modules. Details are provided in the text. <b>FIXME:</b> This figure will have to be replaced once more, as the number of rings in layer 1 of the TBPS increased to 12. . . . .	50
309	4.6	Illustration of the $p_T$ module concept. (a) Correlation of signals in closely-spaced sensors enables rejection of low- $p_T$ particles; the channels shown in green represent the selection window to define an accepted stub. (b) The same transverse momentum corresponds to a larger distance between the two signals at large radii for a given sensor spacing. (c) For the endcap discs, a larger spacing between the sensors is needed to achieve the same discriminating power as in the barrel at the same radius. The acceptance window can therefore be tuned along with a few different values of sensor spacing to achieve the desired $p_T$ filtering in different regions of the detector. . . . .	51
318	4.7	The 2S module (left) and PS module (right) of the Outer Tracker. Shown are views of the assembled modules. The 2S module includes the silicon sensors (yellow), the front-end hybrids hosting 8 chip each (orange), and one service hybrids hosting HV, LV, and the electrical-to-optical converter. The PS module includes the silicon sensors (yellow), the front-end hybrids hosting 8 chip each (orange), one service hybrids hosting the LV and HV, and one hybrid hosting the electrical-to-optical converter. The bottom of the PS module (not visible here) contains the pixellated detector bump-bonded to the readout electronics. Details are given in the text, and in the following sections. . . . .	55
327	4.8	The 2S module (left) and PS module (right) of the Outer Tracker. Shown are details of the module parts. . . . .	56



329	4.9	The 2S module (top) and PS module (bottom) of the Outer Tracker. Shown	
330		are sketches of the front-end hybrid folded assembly and connectivity. The	
331		description of the various electronics components can be found in Sec. 4.4.3. .	57
332	4.10	Electronic system block diagram, exemplified for the 2S module, together with	
333		a labelled sketch of the module. Details are provided in the text. On the data-	
334		trigger-control board (DTC), m-Tx and m-Rx are multi-channel transmit and	
335		receive optical modules. The L1 track-finding block is covered in <b>FIXME: refer</b>	
336		to the trigger chapter . . . . .	58
337	4.11	Illustration of the data flow at the modules' front-end. In the 2S system the	
338		LpGBT is used in 5 G/FEC12 mode (i.e. 6 b at 320 Mb/s from each CIC). In	
339		the PS system the LpGBT is used in 5 G/FEC5 or 10 G/FEC5 mode (i.e. 7 b at	
340		320 Mb/s or 640 Mb/s from each CIC. . . . .	60
341	4.12	Illustration of the CIC data flow and formatting, exemplified for the 2S mod-	
342		ule. . . . .	61
343	4.13	Drawings of a TB2S ladder with its 12 modules (left) and of the support wheel	
344		(right). . . . .	64
345	4.14	Sketch of ladders installed in the support wheel, looking at the wheel from its	
346		end. . . . .	64
347	4.15	Drawing of the innermost layer (layer 1) of the TBPS, showing the central flat	
348		section and the two (identical) tilted sections. The length of each section varies	
349		between the three TBPS layers. The layer 1 also integrates the central part of	
350		the Inner Tracker support tube, as visible inside. . . . .	66
351	4.16	A TBPS layer 1 plank and the layer 1 central flat section with its 18 planks. . .	66
352	4.17	A TBPS layer 1 tilted ring and one of the two layer 1 tilted sections. . . . .	67
353	4.18	The two identical TEDD units, each consisting of five double-discs. Each	
354		double-disc consists of four dees. <b>FIXME: Get picture with modules every-</b>	
355		<b>where? without pink?</b> . . . . .	67
356	4.19	The seven cooling circuits of one dee (left), and a fully assembled dee (right). .	68
357	4.20	A 2S mini-module assembled from a small prototype hybrid comprising two	
358		CBC2 readout chips and two 5 cm long strip sensors, mounted on top of each	
359		other in an aluminium frame. . . . .	69
360	4.21	A full-size 2S module (1.8 mm variant) comprising two flex hybrids with eight	
361		CBC2s each, two 10 cm long strip sensors and aluminium bridges. . . . .	70
362	4.22	Photo of the system test setup with a 2S mini-module (left), and histograms	
363		of the noise of all strips of one CBC2, measured with conventional powering	
364		(blue), a SH placed far away (green) and as close as shown in the photo (yel-	
365		low), plus measurements with data flowing through the VTRx+ (red and grey). .	71

366	4.23	Top: stub reconstruction efficiency of the unirradiated 2S mini-module presented as a function of the stub position. Strip numbers 0 to 126 correspond to the first CBC, strip numbers 127 to 253 belong to the second CBC. Only the region where beam was incident is displayed. The error bars correspond to statistical uncertainties. Bottom: stub reconstruction efficiency for a non-irradiated (red) and an irradiated (blue) 2S mini-module. The mini-module was irradiated to a fluence of $6 \times 10^{14} \text{ n}_{\text{eq}}/\text{cm}^2$ . . . . .	73
367			
368			
369			
370			
371			
372			
373	4.24	Position of services and modules in the material budget model, in cylindrical coordinates and summing over $\varphi$ . Service volumes are represented by black lines. For each module the average sensor position is represented by a single line. Blue lines represent Outer Tracker modules and red lines Inner Tracker modules. . . . .	74
374			
375			
376			
377			
378	4.25	Material budget inside the tracking volume estimated in units of radiation lengths, comparing the Phase-1 detector (left) with the Phase-2 detector (right). The material in front of the Inner Tracker sensors is shown in brown, that inside the Inner Tracker tracking volume is shown in yellow, the material between IT and OT sensors is shown in green and the material inside the Outer Tracker tracking volume is shown in blue. The histograms are stacked. <b>FIXME:</b> Improve label. . . . .	75
379			
380			
381			
382			
383			
384			
385	4.26	Hit occupancy, defined as the fraction of channels containing a digitized hit, as a function of $\eta$ for all layers of the TBPS and TB2S. <b>FIXME:</b> Plot is still for zero pileup. Should include endcaps as well. Improve legend and labels. . . .	76
386			
387			
388	4.27	Stub reconstruction efficiency in TBPS layer 1 as a function of $\eta$ , comparing the flat (solid points) and tilted (open circles) tracker barrel geometries. <b>FIXME:</b> Improve legend and labels, add $p_T$ threshold. . . . .	77
389			
390			
391	4.28	Stub reconstruction efficiency for muons as a function of $p_T$ in the barrel (left) and in the endcap regions (right). <b>FIXME:</b> Improve legend and labels. . . . .	77
392			
393	4.29	Tracking efficiency as a function of the pseudorapidity for single muons with $p_T$ equal to 10 GeV, with 140 pileup events (full circles) and 200 pileup events (open circles). The efficiency is shown for tracks produced less than 3.5 cm from the centre of the beam spot in the radial direction. . . . .	78
394			
395			
396			
397	4.30	Tracking efficiency (left) and fake rate (right) as a function of the pseudorapidity for $t\bar{t}$ events with 140 pileup events (full circles) and 200 pileup events (open circles). The tracks are required to have $p_T > 0.9 \text{ GeV}$ . The efficiency is shown for tracks produced less than 3.5 cm from the centre of the beam spot in the radial direction. . . . .	79
398			
399			
400			
401			
402	4.31	Relative resolution of the transverse momentum (left) and resolution of the transverse impact parameter as a function of the pseudorapidity for the Phase-1 (black dots) and the upgraded (red triangles) tracker, using single isolated muons with a transverse momentum of 10 GeV. . . . .	80
403			
404			
405			
406	4.32	Vertex position resolution in $x$ and $y$ (left) and $z$ (right) as a function of the number of tracks associated to the vertex, for $t\bar{t}$ events with 140 pileup events (full circles) and 200 pileup events (open circles). . . . .	80
407			
408			

409	4.33	Image of a prototype PS-p sensor produced by Novati. Insets show the corner region and macro-pixel punchthrough bias structures and p-stops . . . . .	81
410			
411	4.34	Exploded view of the 1.6 mm (top left), 2.6 mm (top right) and 4.0 mm (bottom) PS module variants <b>FIXME:</b> replace with fig with legend when available. . . . .	84
412			
413			
414	4.35	Images of the PS module jigs used to glue the PS-s sensor to the Al-CF spacers (left) and to glue the MaPSA assembly to the Al-CF spacers (right). . . . .	85
415			
416	4.36	Photo of the PS module jig being used to glue the PS sensor assembly to the CFRP baseplate. . . . .	86
417			
418	4.37	Exploded view of the 1.8 mm (left) and 4.0 mm (right) 2S module variants. <b>FIXME:</b> replace with fig with legend when available. Check TDR! . . . . .	87
419			
420	4.38	Images of the 2S module jigs used to glue the sensors to the Al-CF bridges (left) and to glue the front-end and service hybrids to the sensor package (right). . . . .	88
421			
422	4.39	Left: photo of a prototype setup used to perform the double-sided metrology of bare 2S modules. The bare module is placed on a rotation table. Alignment marks on the module corners of the top and bottom sensor are photographed by the top and bottom camera, respectively. The measurements are referenced to each other by exploiting the fact that the axis of rotation is common. Right: results of a measurement. The four quadrants show zooms onto the corners. The coordinate perpendicular to the strips is denoted as $x$ , while $y$ is the coordinate parallel to the strips. The bottom sensor is shown in blue and the top sensor in pink. The result is reported in the legend; the precision is well within the specifications of $\Delta x < 50 \mu\text{m}$ , $\Delta y < 100 \mu\text{m}$ , rotation angle between strips $< 400 \mu\text{rad}$ . . . . .	89
423			
424			
425			
426			
427			
428			
429			
430			
431			
432			
433	4.40	Cross-section of PS module prototype used for thermal testing. . . . .	89
434	4.41	Electron microscope images of Al-CF spacer before (left) and after (right) O <sub>2</sub> -plasma etching. Images show significant reduction of carbon fibers which appear as darker images in surface of material. . . . .	90
435			
436			
437	4.42	Test setup used to measure leakage current vs. high voltage for Al-CF spacer held between two aluminum blocks. One block was held at ground and measured the leakage current, while voltage was ramped up on the other block. . . . .	90
438			
439			
440	4.43	Leakage current vs. high voltage for Al-CF spacer using test setup described in Fig. 4.42. Two spacers coated with 25 $\mu\text{m}$ Parylene Type N, one with and the other without plasma etching prior to Parylene coated, held voltage up to 1000 V. . . . .	91
441			
442			
443			
444	4.44	Leakage current vs. high voltage for two CFRP baseplates held between two aluminum blocks. One aluminum block was held at ground and measured the leakage current, while voltage was ramped up on the other aluminum block. One of the CF baseplates was coated with 25 $\mu\text{m}$ Parylene Type N, and the other was coated with 25 $\mu\text{m}$ Parylene Type C. For the 1st two tests (labeled "FullCenter"), the aluminum block was larger than the baseplate. For the remaining tests, a small 1" x 1" block was used to measure the leakage current. In all cases voltage held up to 1000 V without breakdown. . . . .	91
445			
446			
447			
448			
449			
450			
451			

452	4.45	Preliminary bunr-in system based on Peltier elements. . . . .	93
453	4.46	Edge view of the PS module showing the position of the MAPSA module in	
454		the assembly. . . . .	94
455	4.47	Image of a dummy MAPSA undergoing tests on an automatic probe station.	
456		The pad pairs alternate between short jumper connections and higher resis-	
457		tance loop connections through a daisy chain. An open region caused by a	
458		scratch has been identified . . . . .	95
459	4.48	Drawing of modules mounted on the plank for layer 1 of the tracker inner flat	
460		barrel. Cooling tubes are bent to mate with neighbor planks in phi. . . . .	97
461	4.49	Carbon fiber/foam support ring for the inner layer of the flat barrel on it's	
462		assembly tooling plate. Locations of the support inserts can be seen in blue.	
463		Fingers projecting to the inner radius are "bumpers" to prevent damage dur-	
464		ing insertion over the pixel support tube. . . . .	98
465	4.50	Thermal model of a module mounted on a plank. The coolant is assumed to	
466		be at -20 degrees C. Hot spots on the edges correspond to the GBT and DC-DC	
467		converter assemblies . . . . .	100
468	4.51	Prototype module mounted on a test plank inside a cold box at SiDet. Multiple	
469		heaters are mounted on the module to simulate heat sources. An array of RTD	
470		temperature sensors are mounted throughout the assembly. . . . .	100
471	4.52	Laser hit efficiency in two neighboring pixels as a function of phase between	
472		the laser pulse and MPA-Light clock (Y-axis) and the coordinate of the laser. . .	102
473	4.53	Noise of the 2S mini module operated at -30 C. The mean and RMS of the	
474		distribution are shown. <b>FIXME:</b> Plot will be updated. . . . .	102
475	4.54	Noise of the three 2S full size module operated at room temperature. Average	
476		noise for each of the 16 CBC's on all 3 modules. The uncertainties are given	
477		by the RMS of the measurement. The noise is derived by fitting a cumulative	
478		distribution function of the normal distribution to the S-curve and the width	
479		is taken as the noise. <b>FIXME:</b> Plot will be updated. Module 3 is missing . . . . .	103
480	4.55	Hit efficiency as a function of the phase difference between the particle ar-	
481		rival time and the 40 MHz clock edge in the MPA. The measured efficiency	
482		spectrum is shown by the black dots, the red curve represents a fit to the mea-	
483		surement, and the black curve shows the jitter corrected efficiency distribution. .	103
484	5.1	Longitudinal segmentation of the HB Phase-1 upgrade detector. Each color	
485		indicates the four sets of layers individually read [33]. . . . .	109
486	6.1	Expected integrated dose in the region of the current CMS calorimeters after	
487		3000 fb <sup>-1</sup> of delivered luminosity. The endcap calorimeters are within the	
488		dashed black line. . . . .	112

489	6.2	Signal loss predictions for the current endcap calorimeter active materials as a function of irradiation dose. (Left) PbWO <sub>4</sub> light transmission spectra for different $\gamma$ and proton doses. All proton doses are lower than the maximum expected by the end of HL-LHC operation. (Right) Fractional decrease of scintillator light in HE layer 1 (closest to the interaction point) as a function of LHC integrated luminosity. . . . .	112
490			
491			
492			
493			
494			
495	6.3	Contribution to EM energy resolution in terms of an effective noise term as a function of $\eta$ for different N(pileup) (= number of overlapping pp interactions) scenarios. . . . .	113
496			
497			
498	6.4	Estimate of the $p_T$ resolution for true forward jets in a VBF Higgs Monte Carlo (MC) simulation as a function of number of layers in the HEC. . . . .	115
499			
500	6.5	Estimate of the $p_T$ resolution for true forward jets in a VBF Higgs MC simulation for different values of intercell calibration precision, assuming a cell size of $\sim 1 \text{ cm}^2$ . . . . .	116
501			
502			
503	6.6	Expected per-cell noise increase as a function of fluence for different cell sizes. . . . .	117
504	6.7	Quarter longitudinal cross section of the HGCal. . . . .	118
505	6.8	Predicted radiation dose and hadron fluence in the HGCal after $3000 \text{ fb}^{-1}$ HL-LHC integrated luminosity. The radius as a function of depth at which the transition is made from silicon to scintillator detector is indicated. <b>Draw in the boundary.</b> . . . . .	120
506			
507			
508			
509	6.9	Drawing of an EE cassette. The cassette shown is for the 14th (last) cassette layer, which includes detector layers 27 and 28. This cassette is approximately 1.3 m long in the radial direction by 1.6 m wide across the outer corners. The upper right inset shows the corner where three modules meet and are secured by a common screw and spring washer. The lower right inset is a cross-section through the thickness of the cassette at the position where the screw secures the modules. <b>Need to update cross-section to be one with motherboards.</b> . . . . .	121
510			
511			
512			
513			
514			
515			
516	6.10	Cassette cooling plate showing the routing of the stainless steel tube that carries the high-pressure two-phase CO <sub>2</sub> coolant. This cooling plate corresponds the layer 12 of the hadronic calorimeter. <b>Need a better version without the extraneous fractions of a mm.</b> . . . . .	122
517			
518			
519			
520	6.11	FH cassette concept. Individual modules are mounted on a copper cooling plate, and flat cables carry services from the module PCB to the outer edge of the cassette . . . . .	122
521			
522			
523	6.12	Module exploded view. <b>Replace with 8" diagram.</b> . . . . .	123
524	6.13	MIP charge collected vs. neutron fluence for six different types of sensors ( $\sim 300 \mu\text{m}$ n-type, $\sim 300 \mu\text{m}$ p-type, $\sim 200 \mu\text{m}$ n-type, $\sim 200 \mu\text{m}$ p-type, and $\sim 100 \mu\text{m}$ n-type, $\sim 100 \mu\text{m}$ p-type) at two different bias voltages (square markers denote 600 V, triangle markers denote 800 V). The $\sim 300 \mu\text{m}$ measurements extend from $4\text{--}9 \times 10^{14} \text{ n/cm}^2$ , the $\sim 200 \mu\text{m}$ measurements extend from $1.5\text{--}4 \times 10^{15} \text{ n/cm}^2$ , and the $\sim 100 \mu\text{m}$ measurements extend from $6\text{--}15 \times 10^{15} \text{ n/cm}^2$ . . . . .	124
525			
526			
527			
528			
529			
530	6.14	Placement of sensors by active thickness in the FH. <b>Replace with 1 MeV n equivalent map.</b> . . . . .	125
531			



532	6.15	Comparison between a plastic scintillator sigma tile (left), currently used in the HE detector and proposed for the low dose regions of BH, and a finger tile (right), the proposed design for the high dose regions. . . . .	127
533			
534			
535	6.16	Diagram (not to scale) of the motherboard-module stackup showing the silicon sensor, module PCB ("Sensor Substrate"), FE ASIC its linear LV regulator on the module PCB, GBT 10-Gbps optical link, concentrator chip, linear LV regulator for the motherboard components, and flexible connection between motherboard and module. . . . .	128
536			
537			
538			
539			
540	6.17	Top (x-y) and side (z-y) views of the motherboard connected to three modules.	129
541	6.18	Comparison of the silicon (left) and scintillator (right) PCB-module stackups. .	129
542	6.19	Diagram of motherboard and ganged readout PCBs covering a single $10^\circ$ wedge of HEC. . . . .	130
543			
544	6.20	Comparison of rates between the Phase 1 electromagnetic object trigger running at an instantaneous luminosity resulting in a mean of 40 pileup interactions per bunch crossing, and the HGCal with a mean of 140 pileup interactions per crossing. . . . .	132
545			
546			
547			
548	6.21	Level-1 single jet trigger rates in the endcaps, comparing the Phase 1 trigger running at an instantaneous luminosity resulting in a mean of 40 pileup interactions per bunch crossing, and the HGCal with a mean of 140 pileup interactions per crossing. . . . .	133
549			
550			
551			
552	6.22	One-MIP signal peak in sensor cells with noise equivalent to 0.3 MIP, after MIP tracking and local isolation algorithms have been applied. The sensor cells are located in the region $2.8 <  \eta  < 2.9$ and have an active thickness of $100\ \mu\text{m}$ . In the simulated events the mean number of interactions per bunch crossing is 140. . . . .	136
553			
554			
555			
556			
557	6.23	Median cell occupancy as a function of layer for signals (top plot) above 0.5 MIP and (bottom plot) above 5 MIP for an instantaneous luminosity such that the mean number of pileup interactions per bunch crossing is 200. . . . .	138
558			
559			
560	6.24	Probability for a sensor pad to be busy during any particular bunch crossing, as function of layer, for different $\eta$ locations, when the instantaneous luminosity is such as to result in a mean of 140 pileup interactions per bunch crossing.	139
561			
562			
563	6.25	Markers: Dose constants of different types of plastic scintillator measured at different dose rates. Dark green lines: necessary dependence of dose constant on dose rate to achieve a given light loss after $3000\ \text{fb}^{-1}$ HL-LHC integrated luminosity. Any scintillator "above" the green lines is sufficient for BH. Teal lines: average dose rates corresponding to a given dose integrated over $3000\ \text{fb}^{-1}$ of HL-LHC operation. . . . .	140
564			
565			
566			
567			
568			
569	7.4	Schematic overview of the CMS DAQ system. . . . .	152
570	8.1	The opened-up CMS detector in the under ground experimental cavern at P5. Cathode Strip Chamber modules (trapezoidal in shape) instrumented with front-end electronics (in aluminum shielding boxes) are clearly visible on the right side. . . . .	160
571			
572			
573			

574	8.2	Schematic of CMS electronics system. . . . .	160
575	8.3	Event Loss Fraction vs Luminosity (the ultimate HL LHC luminosity is $7.5 \times 10^{34} \text{cm}^{-2} \text{s}^{-1}$ ).162	
576	8.4	Measured CFEB (with HL LHC firmware) rate loss vs statistical model. . . . .	163
577	8.5	L1A*LCT rate/luminosity vs luminosity. The L1A*LCT rate is per chamber.	
578		The dashed lines represent the asymptotic rate/luminosity for 2012 and 2015	
579		data. . . . .	164
580	8.6	DCFEB board . . . . .	164
581	8.7	ME2/1 CFEB and DCFEB rate loss curves. . . . .	165
582	8.8	A OTMB mezzanine board. The snap-12 Optical transceiver can be seen on	
583		the left. . . . .	165
584	8.9	ALCT main board with ALCT mezzanine board (red circle) . . . . .	166
585	8.10	Typical GIF++ plot of a muon passing through the 6 planes of the ME1/1b	
586		Chamber. The front axis corresponds to 32 wire strips. Eight time samples	
587		are shown on the axis into the page, and the vertical axis is ADC counts. Col-	
588		ors emphasis ADC counts (measured charge) in a given time sample. Min-	
589		imum ionizing particles have typically light brown level deposits, while red	
590		and black correspond to greater than 3 and 5 times minimum ionizing respec-	
591		tively. . . . .	168
592	1	Space allocation in the tracking volume. <b>FIXME:</b> The naming in the drawing	
593		needs to be updated: change Pixels to Inner Tracker. . . . .	179
594	2	Left: photo of the currently installed services on the CMS central wheel, YB0.	
595		Right: photo high-lighting in blue one of the tracker service channels and	
596		PP1s, and in pink one of the neighbouring ECAL+HCAL service channels. . .	181
597	3	The cooling plants of the pixel Phase-1 CO <sub>2</sub> cooling system at CERN. A similar	
598		system, but larger in scale, will be used for the Phase-2 tracker. . . . .	183

DRAFT



# Tables

600	3.1	The expected number of events, after the analysis selections, for same-sign WW and WZ final states. Signal, irreducible background (I-bkg) and reducible one (R-bkg) are reported. Each line corresponds to a possible scenario for the CMS operations. . . . .	39
601			
602			
603			
604	3.2	Science Goals → Science Requirements Matrix . . . . .	42
605	3.3	Science Requirements → Technical Requirements Matrix . . . . .	43
606	4.1	Most important specifications of the three sensor types used in the Outer Tracker. <b>FIXME:</b> these numbers have to be double checked. . . . .	52
607			
608	4.2	Main parameters of the 2S module and the PS module. . . . .	54
609	4.3	Summary of module types and versions in the Outer Tracker. <b>FIXME:</b> Does not yet include the change to twelve rings in TBPS layer 1. . . . .	55
610			
611	4.4	Summary of expected thermal runaway turn-on temperatures for the various module types and sub-detectors. Calculations were performed using the expected nominal heat loads after $3000 \text{ fb}^{-1}$ of integrated luminosity. <b>FIXME:</b> Numbers for PS modules to be added when available. . . . .	58
612			
613			
614			
615	4.5	Summary of responsibilities for module assembly parts and assembly fixtures. <b>FIXME:</b> how many MAPSA for bump-bonding? Check with Ron's section! . . . .	82
616			
617	4.6	Geometry of the flat barrel section. R values are in mm. . . . .	97
618	4.7	Results of prototype mechanical tests . . . . .	100
619	4.8	Summary of MaPSA-Light assemblies production and quality. ( $A1=Aemtec$ , $A2=Novapack$ , $A3=Leti$ , $A4=CVI$ ) . . . . .	101
620			
621	5.1	Characteristics of plastic scintillators and WLS fibers considered for usage in the HB upgrade; OD indicates over-doped scintillators. . . . .	108
622			
623	6.1	FH silicon sensor arrangement: thickness of active silicon layer with the associated cell size and $S/N$ for a MIP before and after an integrated luminosity of $3000 \text{ fb}^{-1}$ . . . . .	125
624			
625			
626	6.2	Maximum bandwidth per module for trigger and readout data as compared to current allocation of optical or electrical links emanating from the concentrator chip for the FH modules. "o-link" refers to optical links and "e-link" refers to Twinax electrical links. . . . .	131
627			
628			
629			

630	7.1	Number of input channels and boards for parallel calorimeter trigger processing	148
631	7.2	Optical link inputs from the muon detectors to the Muon Track-Finder. . . . .	149
632	7.3	DAQ/HLT system parameters. . . . .	152
633	1	Quantities and estimated dimensions of the Inner Tracker and Outer Tracker	
634		services running from the bulkheads to the PP1s, for both ends of the tracker.	
635		<b>FIXME:</b> Check and complete the table. . . . .	181
636	A.1	Per-layer counts of modules of various types in the endcap calorimeter . . . . .	185

DRAFT

637 **Chapter 1**

638 **Executive Summary - Editor O'Dell 2 pages**

639

- Introduction: Short motivation on physics and international context
- Scope: Brief on international scope, then scope of the US project (i.e paragraph or two on each of the tracker, calorimeter, triggerDAQ, laying out the US scope and the international context)
- Cost and Schedule (1 par.)
- Acquisition Strategy (1 par.)

DRAFT

DRAFT

## Chapter 2

# Project Overview - Editor O'Dell, Hill, Snow, Spalding 30 pages

## 2.1 Overview of the HL-LHC Physics Program

Since its inception, the physics program of the CMS experiment has sought to address important open issues in high energy physics such as the origin of elementary particle mass. Significant progress on this question was made during the first major run of the LHC when a new particle with mass  $\sim 125$  GeV was observed by the ATLAS and CMS collaborations [1, 2]. The new particle appears to be the quanta of the Higgs field that imbues elementary particles with mass depending on the strength of their interaction with it. Additional data collected in Run I, and now Run II, of the LHC have allowed increasingly precise studies of the properties of the newly observed particle that test this interpretation. For example, by measuring how often the new boson interacts with each of the other fundamental particles, i.e. their “coupling” to the Higgs, the data analyzed already shows that these couplings seem to have the mass dependence uniquely characteristic of the Higgs field.

The uncertainties on these measurements are large enough, however, that it remains to be seen if the Higgs boson that we have observed is solely responsible for fundamental particle mass as predicted by the Standard Model (SM), or whether there are other Higgs bosons that also play a role. Moreover, since the Higgs boson that we have observed appears to couple to all known fundamental particles, it is reasonable to assume that it would couple to any new beyond the Standard Model (BSM) particles. In fact, for the one type of BSM particle that we know exists from astrophysical and cosmological observations, i.e. those composing dark matter (DM), it is possible that the Higgs is the only SM to which it couples via what is called a “Higgs portal” to DM. Both of these issues, whether or not there are additional Higgses and/or BSM particles that couple to the Higgs, can be addressed by making two types of measurements are important. How often the Higgs is produced in a specific way tells us the coupling with certain particles and, once produced how often the Higgs decays in a specific way tells us the coupling with other particles. By combining this information, if we get enough Higgs interactions of each type (and have an experimental apparatus capable of distinguishing them) we can obtain a complete picture of the role of the Higgs in generating mass. To get enough interactions of each type we will need substantially more data than the  $300 \text{ fb}^{-1}$  the LHC was designed to provide. To meet this physics goal, the accelerator will be upgraded to the high luminosity LHC (HL-LHC) that will produce  $10\times$  the amount of data,  $3 \text{ ab}^{-1}$ . And to have an experimental apparatus capable of dealing distinguishing the various production/decay modes in the face of complications introduced by the high intensity beams of the HL-LHC, the CMS detector must be likewise upgraded. Figure 2.1 shows the expected improvement in the precision of the Higgs boson coupling measurements from the current CMS results (left) to the precision that

will be obtained with a dataset of  $3 \text{ ab}^{-1}$  recorded by an upgraded CMS detector [3, 4].

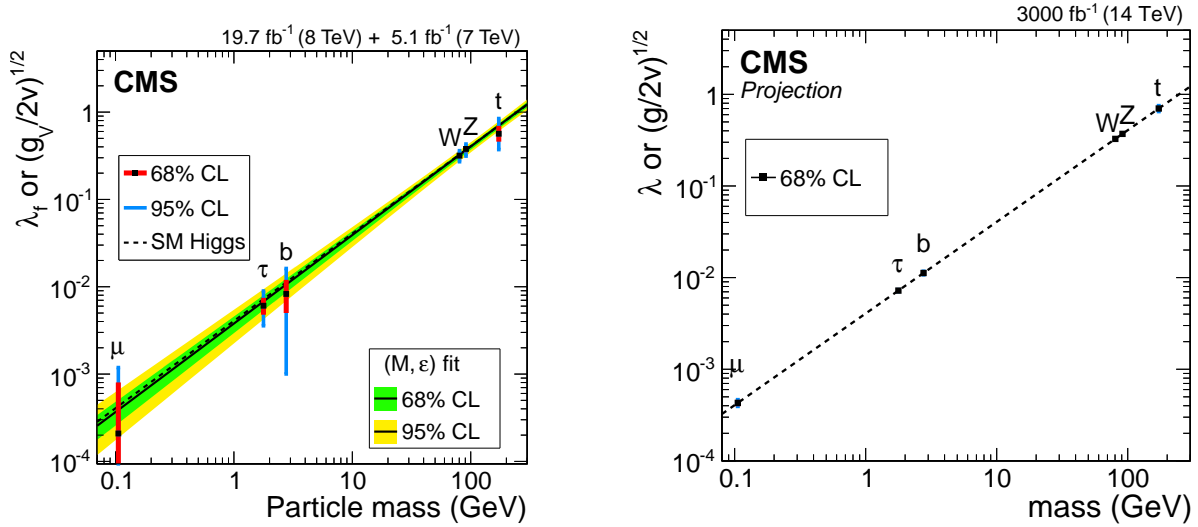


Figure 2.1: Observed and projected precision on Higgs boson couplings as a function of boson or fermion masses.

Despite the progress in understanding the origin of mass (and in some cases because of it), there are other important questions that remain that we will also address with the data from the HL-LHC. For example, the scalar nature of the Higgs particle presents new challenges since in quantum field theory such particles receive very large quantum corrections to their mass. This problem, referred to as the hierarchy problem, implies that either new particles must appear in the LHC data with masses  $\sim 1 \text{ TeV}$  to cancel these divergent contributions, or the Universe's fundamental parameters have to be extremely "fine-tuned" in order to produce the observed mass value of  $125 \text{ GeV}$ . A principal goal of the ongoing LHC physics program is perform a sufficiently comprehensive search for these new particles so that they are either discovered, or all such "natural" solutions to the hierarchy problem are conclusively excluded so that the fine-tuned condition of the Universe can thereby be deduced. Either outcome would be a profound statement about Nature.

A related goal of the ongoing LHC physics program is to perform a comprehensive search for BSM physics related to the galactic DM referred to above. Depending on the particle nature of DM, it may couple to SM particles and could therefore be directly produced by LHC collisions. Because DM is at best weakly interacting, the production of such particles can only be inferred through the presence of missing transverse energy (MET) via so called "mono-X" analyses. These searches turn out to be competitive with the direct search experiments in the appropriate regions of comparison, and projections indicate that the HL-LHC can drive this search below the neutrino coherent scattering limit, which will be a concern for the direct experiments [5].

In carrying out this search program, the pp data from Run I, and now Run II, have been thoroughly analyzed looking for evidence of various forms of BSM physics that would provide a natural solution to the hierarchy problem and/or DM. No such evidence has yet been found in any of these analyses and ever more restrictive constraints are placed on the various BSM models. For example, Figure 2.2 summarizes the 95 % CL exclusions of the predicted new states in one particularly well motivated BSM scenario, supersymmetry (SUSY). While the constraints derived thus far significantly exclude particles produced through the strong interaction, the production of BSM particles via rarer mechanisms such as electroweak interactions or



## 2.2 Overview of the CMS Experiment and the Upgrades needed for HL-LHC

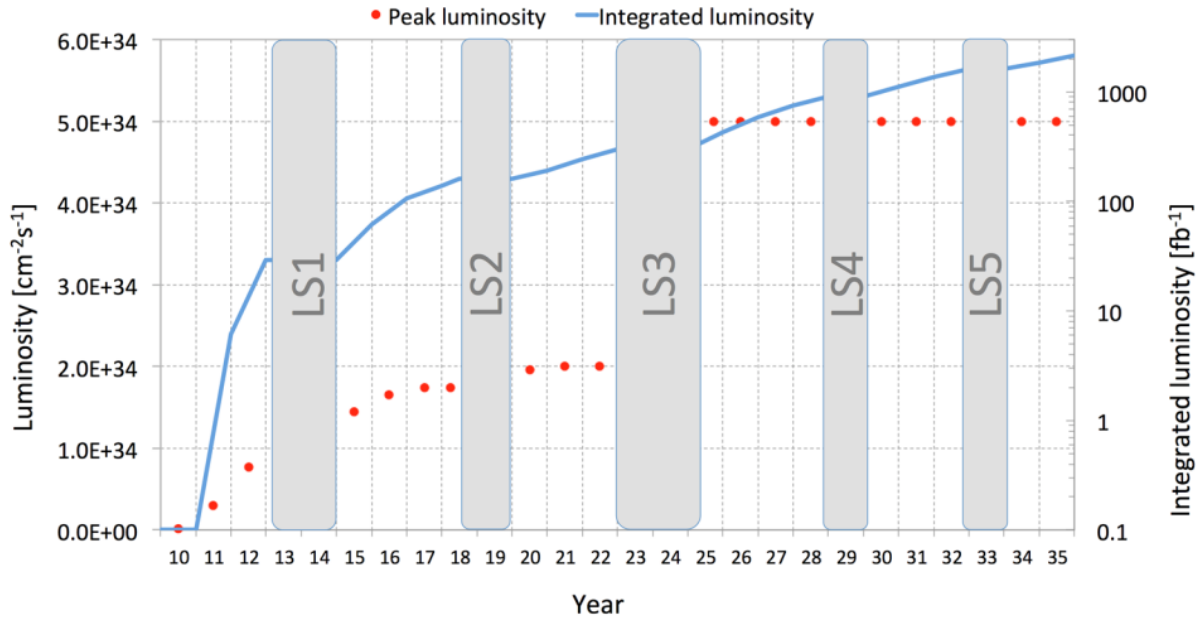


Figure 2.3: Projected LHC performance through 2035, showing preliminary dates for long shutdowns of LHC and projected luminosities.

CERN began planning to increase the luminosity of the LHC above the original design even before the machine went into operation. Major revisions to the machine or the experiments require access to the accelerator tunnels and the experimental areas that can only be accomplished efficiently during long shutdown periods. The current plan calls for a series of long periods of data-taking, referred to as Run-I, Run-II, etc. interleaved with long shutdowns, designated LS1, LS2, LS3.

Run-I is the name given to the completed data-taking period in 2011 and 2012. During the first long shutdown, LS1, which started in 2013 and ended at the beginning of 2015, modifications were made to the LHC to enable it to run at the center-of-mass energy of 13 TeV. The energy will be raised closer to the design energy of 14 TeV over time.

Run-II began in 2015. For this run, the bunch spacing has been reduced to 25 ns from 50 ns. With this bunch spacing, the original performance goal for the LHC, to operate at an instantaneous luminosity of  $1 \times 10^{34} \text{cm}^{-2}\text{s}^{-1}$  is likely to be achieved early in Run-II. Under these conditions, CMS will experience an average of about 25 inelastic interactions per bunch crossing, referred to as event pileup (PU). This is the operating scenario for which the CMS experiment was originally designed. With several planned improvements to the accelerator chain, it is expected that the peak luminosity will reach  $2 \times 10^{34} \text{cm}^{-2}\text{s}^{-1}$ , providing an integrated luminosity of over  $300 \text{fb}^{-1}$  by 2023.

By 2023, the quadrupoles that focus the beams at the ATLAS and CMS collision regions will be close to the end of their lives due to radiation exposure. There will a long shutdown, LS3, to replace these with new low- $\beta$  quadrupole triplets. In addition, crab-cavities will be added to optimize the bunch overlap at the interaction region. These changes will produce a significant increase in instantaneous luminosity. Consequently the period of operation that will follow LS3



is called HL-LHC. The proposed operating scenario for HL-LHC is to level the instantaneous luminosity at  $5 \times 10^{34} \text{cm}^{-2}\text{s}^{-1}$  from a potential peak value of  $2 \times 10^{35} \text{cm}^{-2}\text{s}^{-1}$  at the beginning of fills, and to deliver  $250 \text{fb}^{-1}$  per year for a further 10 years of operation. Under these conditions the event PU will rise substantially to become a major challenge for the experiments, and the performance degradation due to integrated radiation dose will need to be addressed.

The schedule of beam operations and long shutdowns, together with projections of the peak and integrated luminosities, is shown in Fig. 2.3. The basic goal of the HL-LHC upgrade is to maintain the excellent performance of the CMS detector in terms of efficiency, resolution, and background rejection for all the physics objects used in the analysis of the data. The main challenges that must be overcome to achieve this goal are radiation damage to the CMS detector from the high integrated luminosity of the HL-LHC and the very high “pileup” that comes from the high instantaneous luminosity. In the following subsections, each of these challenges is described in general terms.

### 2.2.1 Radiation Damage to the CMS Detector at the HL-LHC

The LHC will produce collisions at a rate of about  $5 \times 10^9/\text{s}$ . The particles emerging from these collisions and the radioactivity they induce in the material of the detectors and the on-board electronics will cause significant damage and could result in a progressive degradation of the detector performance. Maintaining the detector performance in the harsh conditions of the LHC was a major consideration in the initial design of CMS and of the Phase-I upgrade. When one considers that the annual dose delivered to the detector per year in the HL-LHC era will be similar to the total dose of all operations from the beginning of the LHC program to the start of LS3, the magnitude of the problem becomes clear.

The main source of radiation is from the particles produced in the proton-proton collisions. The charged particles, mainly pions, produce ionization in the detectors they pass through. They also undergo nuclear interactions that produce cascades of particles that add to the radiation load. Photons, mainly from  $\pi^0$  decays, interact in the material of the beam pipe or tracking systems to form  $e^+e^-$  pairs or reach the calorimeters where they produce electromagnetic cascades. Particles are also backscattered from the calorimeters or escape from cascades within them. These particles spread out and interact with other detector components. Neutrons, in particular, may travel long distances, slowing down and scattering many times in the CMS detector, which is largely hermetic. When the neutrons interact, they can also produce photons and electrons. This results in a mixed field of very low energy neutrons, photons, and electrons that have lost any correlation with the bunch structure of the original collisions and form a relatively uniform background in space and time within the detector volume. Since personnel will need to work on and around the detector when the beam is off, simulations are also performed to predict material activation and dose to personnel.

In order to design a detector that will maintain its performance at the HL-LHC and continue to perform well as the integrated luminosity approaches  $3 \text{ab}^{-1}$ , predictions of the dose rate and particle fluence for each type of particle is needed. Simulations are used to predict the magnitude and composition of radiation as a function of luminosity. The radiation simulations are performed with Monte Carlo transport codes MARS’109 [6] and FLUKA 2011.2b.6 [7] [8]. A description of the codes, their functionality, the CMS geometry models and settings are given in Appendix ???. The impact of each type of particle on the various candidates for material used in the detector must also be known. Information on the performance of the current detector is obtained from test beam measurements, special radiation exposures, and from Run-I experience. The beginnings of radiation damage are clearly observed in Run-I and the results are

used to benchmark the simulations.

For most simulations, a p-p collision is used as the primary event, except for machine-induced background (MIB) simulations, which are described in Appendix ?? . The event generator DPMJET III [9] (version 3.0-6) is used to create the primary proton-proton events in radiation simulations. It is directly linked to the FLUKA code and used as the default event generator for high energy hadronic interactions. The pseudorapidity and transverse momentum distributions are shown in Fig. 2.4. While the multiplicity decreases with increasing  $|\eta|$ , the mean momentum increases rapidly.

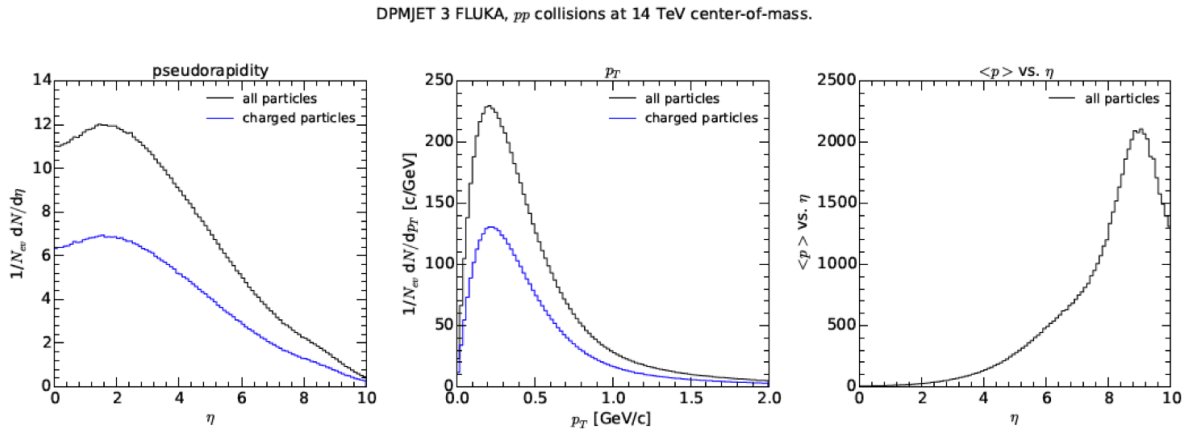


Figure 2.4: Pseudorapidity and transverse momentum distributions of the particles generated by the DPMJET III generator. The black line is for all particles, the blue line is for charged particles.

The normalization used for the prediction of dose and fluence depends on the inelastic collision rate and the duration of LHC operation (total radiation period). For an instantaneous luminosity of  $5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ , and an inelastic cross section of  $\sim 80 \text{ mb}$ , an average of  $4 \times 10^9$  inelastic p-p events per second are produced<sup>1</sup>. For consistency with previous reports, cumulative estimates of particle fluences and dose are based on a previous LHC schedule, where it was assumed that the CMS experiment will receive  $3000 \text{ fb}^{-1}$  by 2035. However, estimates of the residual activity, where the time structure of LHC operation is significant, are based on a pp collision time profile with a total of  $2827.5 \text{ fb}^{-1}$  until 2035 [11].

Two main FLUKA CMS geometries are used to perform simulations for HL-LHC conditions; the “TP Baseline” geometry, derived from the latest nominal Phase-I CMS geometry [12] with the presently installed electromagnetic and hadron endcap calorimeters; and a modified geometry that includes specific elements to represent the Phase-II detector upgrades. A full description of both geometries, with corresponding diagrams can be found in Appendix ??.

An example of the predictions of expected radiation levels for HL-LHC conditions in the CMS detector is given in Figure 2.5, which shows the distribution of absorbed dose over the CMS detector for an integrated luminosity of  $3000 \text{ fb}^{-1}$ .

The damage produced in the detectors by this radiation varies from sub-detector to subdetector. For silicon detectors, radiation produces defects in the silicon lattice that change the bulk

<sup>1</sup>The inelastic non-diffractive cross section of  $\sqrt{s} = 14 \text{ TeV}$  protons is predicted by various event generators. Being the only event generator tuned with LHC data, the results of EPOS LHC is used, which predicts  $80 \text{ mb}$  inelastic cross section [10].

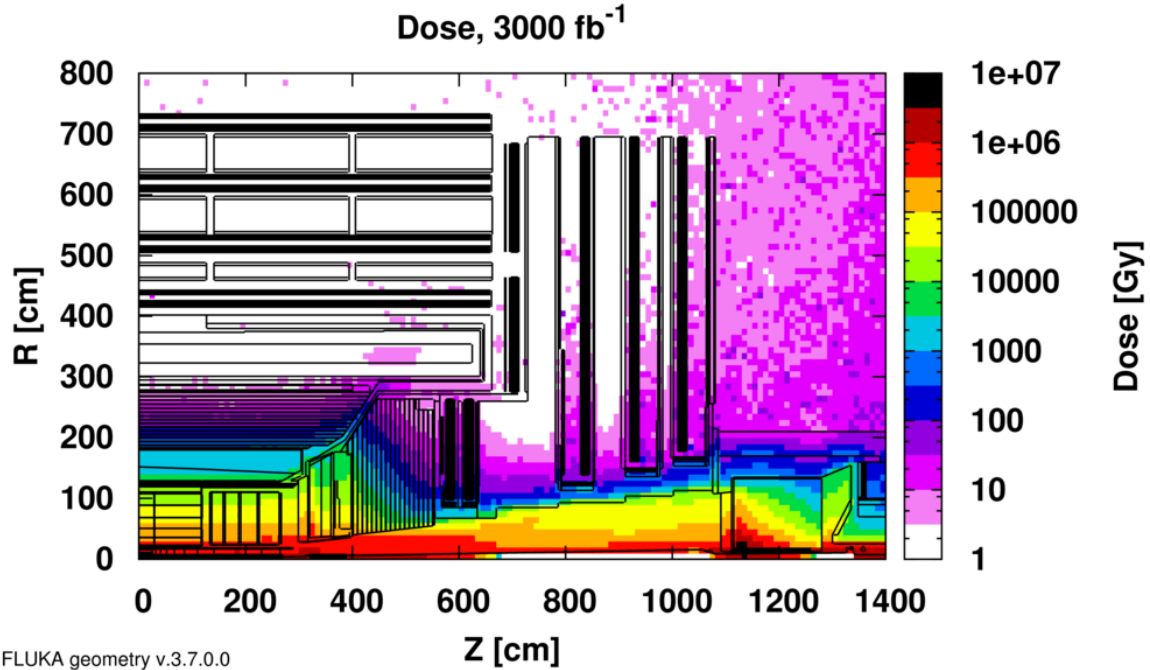


Figure 2.5: Absorbed dose in the CMS cavern after an integrated luminosity of 3000 fb<sup>-1</sup>. R is the transverse distance from the beamline and Z is the distance along the beamline from the Interaction Point at Z=0.

electrical properties of the silicon. One consequence is that leakage currents will increase. Trapping centers for charge carriers reduce the Charge Collection Efficiency (CCE), which in turn lowers the signals from charged particles. The full depletion voltage (the voltage required to make the full thickness of the detector depleted from charges, hence making it fully sensitive to a crossing ionizing particle) increases, eventually reaching unsustainable levels and forcing operation with only partial depletion, resulting in lower signals. At the HL-LHC, some detectors will be subjected to fluences as high as  $10^{16}$  particles/cm<sup>2</sup>, which will make efficient operation difficult. For calorimeters, which in CMS are mainly scintillating  $PbWO_4$  crystals or plastic scintillating tiles with wavelength-shifting fibers embedded in them, the main problem is the loss of transmission of the media through which the scintillation light or wavelength-shifted light must pass. The actual scintillation method does not appear to be harmed by the radiation. This results in a reduction in the signal that can be quite large, in some cases more than 90%, and a corresponding reduction in the resolution. Strategies for mitigating the effects of radiation vary from sub-detector to sub-detector and are described in subsequent chapters.

## 2.2.2 High Pileup

Each of the colliding beams at the LHC consists of many intense “bunches” of protons. Each bunch has a length with rms of  $\sim 5$  cm, transverse dimensions of about 10 microns, and contains a few  $\times 10^{11}$  protons. Bunches will be separated in time by 25ns, corresponding to a spatial separation of approximately 750 cm. There are  $\sim 2800$  filled bunches in each beam and this number cannot be substantially increased. The collision of two bunches is called a “bunch crossing” or “BX” and these occur at a rate of 40MHz. At the nominal luminosity of the HL-LHC, the average number of interactions in a single crossing is approximately 140. Most of these interactions are “soft” or “peripheral” collisions that, if not well understood, are at least well-characterized and do not contribute to the search for new physics at the 0.1-few TeV scale.

As shown clearly in Fig. 2.4, they contain low  $P_T$  particles and produce little energy in the CMS detector. A relatively small fraction of all collisions are “hard” collisions that contain high transverse momentum particles that may come from new high mass objects. Nevertheless, the presence of some tracks and energy from 140 (on average) extra collisions can confuse or degrade the triggers and the offline reconstruction of the hard scatter.

There are, in fact, three kinds of pileups based on the time at which energy is deposited in the tracking detectors to form hits and in the calorimeters to form “showers”. In-time pileup (IT) refers to “hits” or energy deposits from the “extra” p-p collisions in the current bunch-crossing other than those from the collision containing the hard scatter of interest. This is the largest source of hits in the tracking system and also produces significant energy deposition in the calorimeters. In addition, there is out-of-time pileup (OOT), which comes in two varieties: early out-of-time pileup, which refers to energy left in calorimeters from previous BXs in the crossing of interest; and late out-of-time pileup, which refers to energy from later BXs that is integrated along with the trailing portion of the pulse from the BX of interest.

Since the number of bunches cannot be increased, luminosity increases at the LHC result in higher pileup. Pileup produces many more hits in the tracking detectors, leading to mismeasured or misidentified tracks. It also adds extra energy to the calorimeter measurements, such as jet energies, associated with the collision that contained a hard scatter. Electroweak phenomena, which are of special interest, are often characterized by having “isolated” leptons, that is leptons or photons with very little activity around them. Energy or tracks from pileup can contribute to an activity that is not due to the collision containing the leptons or photons and cause them to appear non-isolated. Pileup confuses the trigger and also the offline reconstruction and interpretation of events. It increases the amount of data that has to be read out in each BX that contains a hard scatter. In fact, at the HL-LHC, most of the data read out will be associated with the “pile-up” collisions rather than the collision containing hard scatters. It also increases the execution time for the reconstruction of events in the High Level Trigger and the offline analysis.

In-time pileup can be observed in a single bunch-crossing by the many collision vertices that are reconstructed by the tracking system. A relatively high-pileup crossing that was produced in a special data run in 2012 is shown in Fig. 2.6. There are 78 reconstructed vertices. The total number of pileup collisions is actually somewhat larger because some vertices have too few tracks to be reconstructed. The upgraded tracking system can be designed with enough additional segmentation to associate charged particles with the correct interaction vertices most of the time, even for PU of 140 or 200. This enables the collision containing the hard scatter to be correctly reconstructed and for isolated leptons to be correctly identified in most cases.

The calorimeters in CMS do not have “pointing capability” so it is not possible to associate showers in them with particular vertices. However, the particle flow techniques that are now employed will associate charged tracks in the shower with a particular vertex and this helps to arrive at the correct interpretation of events even in the presence of very high pileup. Moreover, the many simultaneous “typical” or “minimum bias” collisions in the bunch crossing produce a rather smooth energy distribution that can be extrapolated into jet cones or isolation cones to enable a subtraction of their effect.

Out-of-time pileup is illustrated in Fig 2.7. The degree of OOT depends on the intrinsic time spread and jitter of the pulses produced in each detector by particles passing through it and by shaping times and other characteristics of the readout electronics. Tracking systems typically respond in times short compared to the inter-bunch spacing of 25 ns and are not very sensitive to OOT. Calorimeters may produce longer signals and may need longer shaping times



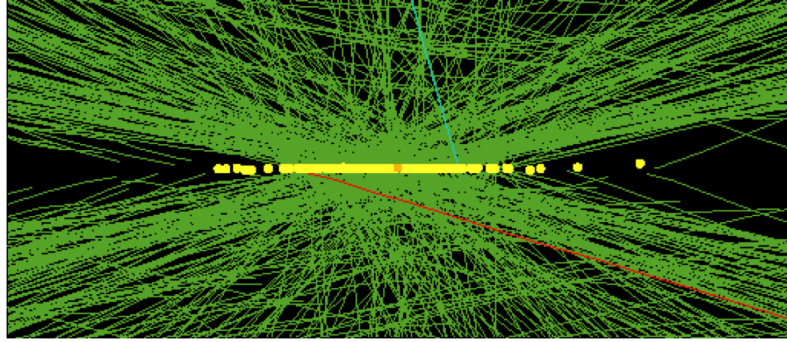


Figure 2.6: High pileup event with 78 reconstructed vertices taken in 2012

and so may suffer from problems with OOT. If the detectors and their readout electronics provide timing or pulse shape information, it is possible to use it to correct the energy deposition associated with a bunch crossing for the energy leakage into that crossing from OOT. Timing measurements are likely to play a more significant role in the HL-LHC era than they have in Run-I, which had 50 ns bunch spacing, in order to cope with out-of-time pileup.

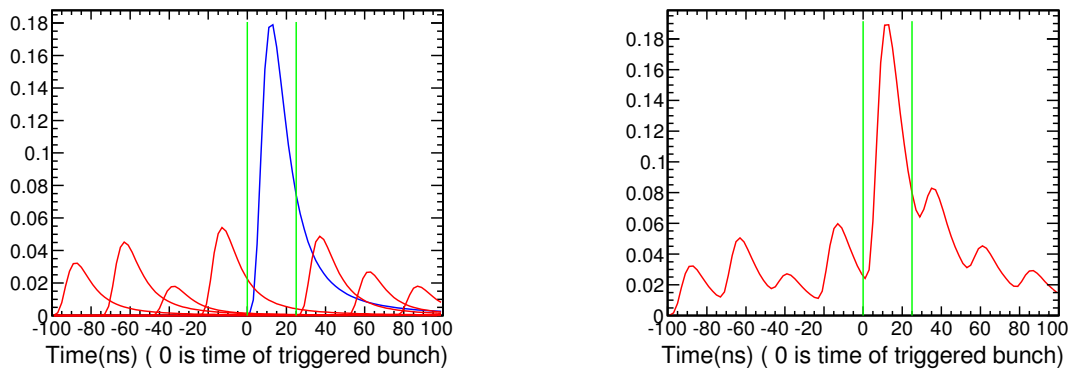


Figure 2.7: Illustration of out-of-time pileup: (left) Individual pulses from a detector as a function of time in ns. The “triggered” or “signal” BX is located at “0” and is bounded by the green vertical lines. The “signal” pulse appears in blue and extends into the next two 25ns “buckets”, which contain some pulse height from the later pileup collisions. Some energy from the preceding “bucket” also falls into the “signal” BX. (right) The sum of the pulse heights as a function of time. The challenge is to correct the “summed” pulse height or charge (related to area under the pulse height distribution) to derive the pulse height or the total charge in the blue signal pulse.

### 2.2.3 A brief introduction to the CMS detector

An exploded view of the CMS detector is shown in Figure 2.8. At the heart of the experiment is a 13 m long, 6 m diameter, 4T superconducting solenoid providing large bending power for momentum measurements and whose return field is large enough to saturate the iron plates in the return yoke, enabling it to be used for muon momentum reconstruction. The gaps between the plates provide slots for four muon tracking stations, each of which consists of several layers of aluminum drift tubes (DT) in the barrel region and cathode strip chambers (CSCs) in the endcap region. Each system is complemented by resistive plate chambers (RPCs).

The bore of the magnet is large enough to accommodate the tracking and calorimetry systems. The tracking volume is contained in a cylinder of 5.8 m length and 2.6 m in diameter. CMS employs ten layers of silicon microstrip detectors, which provide the required granularity and precision to reconstruct efficiently charged tracks in high multiplicity events. The silicon microstrip tracker with its long bending path, combined with the strong solenoidal field, provides excellent momentum resolution. In addition three layers of silicon pixel detectors in the barrel region, complemented by two forward disks at each end, seed track reconstruction and improve impact parameter measurements, as well as providing points with sufficient resolution to reconstruct secondary vertices from decays of particles containing  $b$  and  $c$  quarks.

The electromagnetic calorimeter (ECAL) provides coverage up to pseudorapidity  $|\eta| = 3$  and uses blocks of lead tungstate crystals whose scintillation light is detected by silicon avalanche photodiodes (APDs) in the barrel and vacuum phototriodes (VPTs) in the endcaps. A preshower system is installed in front of the endcap ECAL for  $\pi^0$  rejection. The ECAL is surrounded by a brass/scintillator sampling hadron calorimeter (HCAL) with coverage up to  $|\eta| = 3$ . The light is converted by wavelength shifting (WLS) fibres embedded in the scintillator tiles and channeled via clear fibres for readout to hybrid photodiodes (HPDs) that can operate in high axial magnetic fields. This central calorimetry is complemented by a “tail-catcher (HO)” in the barrel region insuring that hadronic showers are sampled with nearly eleven interaction lengths. Coverage up to  $|\eta| = 5$  is provided by an iron/quartz-fibre calorimeter (HF). The Cherenkov light emitted in the quartz fibres is detected by photomultipliers. The HF ensures nearly full geometric coverage for measurement of the transverse energy in the event.

CMS is triggered by dedicated custom electronics in the Underground Control Room (USC55) which form various partial triggers using trigger primitives from the front ends of the calorimeters and muon detectors. These are then sent to the Global Level 1 trigger which is designed to handle up to 100 kHz rate; the latency is  $3.6 \mu\text{s}$ . Data must be stored on detectors during Level 1 processing. When a Level 1 accept occurs, data fragments from individual detectors are sent to the High Level Trigger (HLT), operating on a large computer cluster, to build complete events. The HLT performs a lean version of the offline reconstruction using full event data and uses the result to decide if the event should be written, together with trigger information, to mass storage for subsequent analysis. In Run-I CMS wrote out about 500-1000 events/second. A detailed description of the CMS detector is given in reference [13].

Since the LHC luminosity will exceed the original luminosity of  $10^{34} \text{cm}^{-2} \text{s}^{-1}$  for which CMS was designed, there is an ongoing upgrade that will allow CMS to perform well until the start of the HL-LHC. Modifications to the detector described above include the completion of the fourth layer of endcap muon detectors, which was only partially implemented in Run-I; improvement in the muon electronics; the replacement of the 3 barrel layer, 2 endcap disk pixel detector with a 4 barrel-layer, 3 endcap disk pixel detector; the provision of longitudinal segmentation in the barrel and endcap hadron calorimeters; replacement of the HPDs for the barrel and endcap hadron calorimeters with Silicon Photomultipliers (SiPMs); the replacement of single anode photomultipliers in the HF with multianode photomultipliers; and the modernization and upgrade of the trigger and data acquisition systems to handle higher data volumes.

## 2.2.4 General Considerations for the Phase-II upgrade

For the HL-LHC, the brightness of beams and the new focusing/crossing scheme at the interaction point will enable the accelerator to potentially deliver a luminosity of  $2 \times 10^{35} \text{cm}^{-2} \text{s}^{-1}$  at the beginning of each fill. This would increase the interaction rate and collision PU beyond the capabilities of existing and envisioned detector and trigger technologies. It is therefore

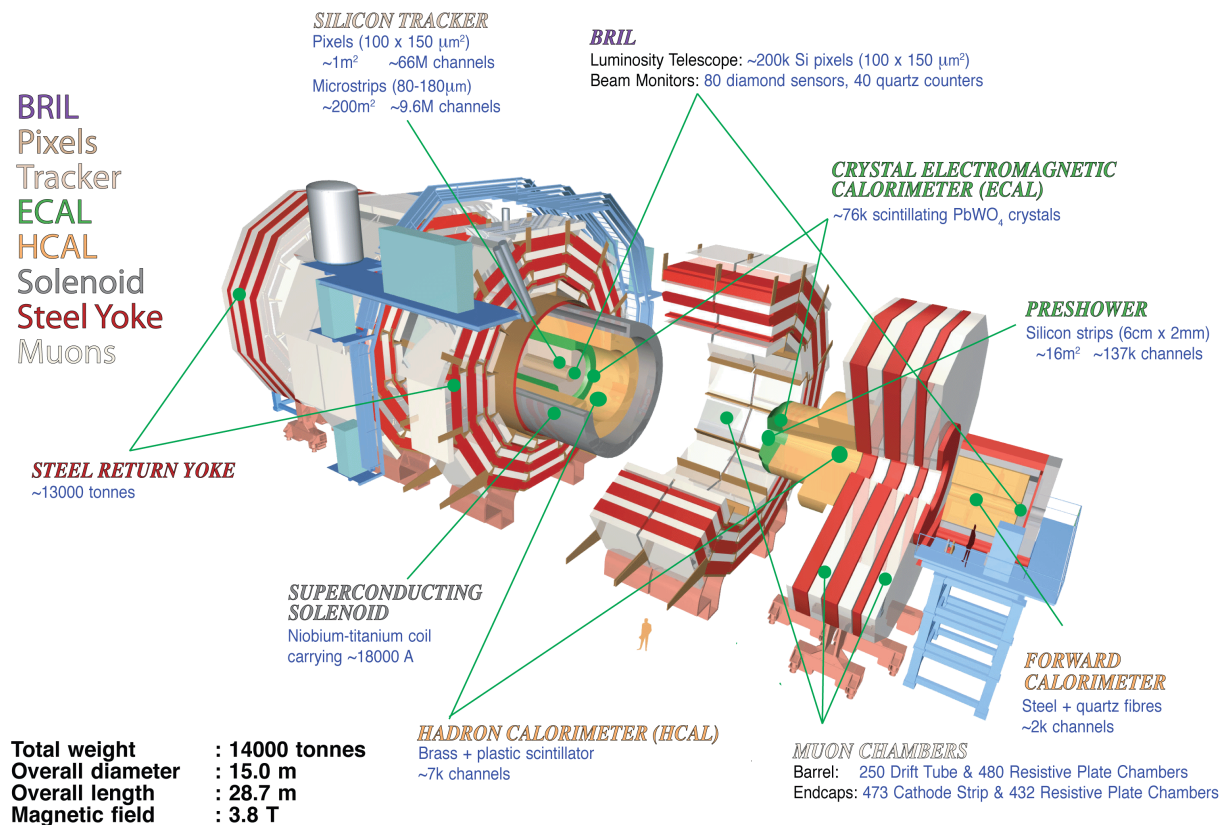


Figure 2.8: A schematic representation of the CMS Detector in Run 1, with its various sections in retracted positions. The central yoke block is YB0. The first block (YB+1, and corresponding YB-1) is shown partially moved away from YB0. The second yoke block (YB+2, which has a corresponding YB-2) is shown fully moved past the solenoid vacuum tank. The end-cap calorimeters are attached to the first endcap disk YE+1, then endcap CSCs and RPCs, then YE+2, more muon chambers, and YE+3, with additional muon chambers on the front and back. Eventually, another disk, YE+4 will be added at the end to provide shielding from beam related backgrounds. This configuration is repeated on the other end. In operation, the detector is closed by moving all pieces together.

proposed to maintain a lower, but stable instantaneous luminosity by continuously tuning the beam focus and crossing profile throughout the duration of beam fills in a process referred to as luminosity levelling. The nominal scenario is to operate at a leveled luminosity of  $5 \times 10^{34} \text{cm}^{-2} \text{s}^{-1}$ , corresponding to a mean pileup of 140 interactions per beam crossing<sup>2</sup>. The primary goal of the Phase-II upgrade program is therefore to maintain the excellent performance of the Phase-I detector under these challenging conditions throughout the extended operation of HL-LHC.

A major focus of CMS has been to identify changes that are mandatory for the beam conditions of HL-LHC and significant effort has been expended to understand the effect of radiation damage. Performance projections are based on a combination of detailed measurements using the data taken in the experiment throughout the period 2011-2012 and the exposure of test components to radiation levels matching anticipated HL-LHC doses. From these studies it is very clear that the tracker and the endcap calorimeters must be replaced for Phase-II.

<sup>2</sup>An average of 140 p-p collisions per beam bunch crossing assumes a high value of the total interaction cross section extrapolation to 14 TeV, and takes into account bunch to bunch fluctuations as discussed in ref [14]

With these required changes, the performance issues associated with high PU, that are also the most pronounced in the inner and forward detector regions, can be addressed. Pile-up mitigation in CMS heavily relies upon particle-flow event reconstruction. To this end, the tracker granularity can be increased to maintain the excellent tracking efficiency to enable the determination of the original p-p collision points for all charged particles. New endcap calorimeter configurations will also provide the opportunity to optimize segmentation and improve energy resolution, particularly for jets.

The ability to ensure efficient event selection for data acquisition is a key prerequisite to fully benefit from increased luminosity. The precise study of the relatively low-mass Higgs boson discovered in 2012, and the search for new particles occurring in cascade decays will require continued use of low transverse momentum,  $p_T$ , trigger thresholds. To achieve this, the trigger electronics (i.e. the L1 trigger) must be upgraded. A sufficient reduction in trigger rate can only be accomplished by improving  $p_T$  resolution to obtain lower rates without loss of efficiency, and by mitigating the effect of the combinatorial backgrounds arising from PU. A new approach is therefore required, namely the introduction of tracking information at L1, providing the capability to implement trigger algorithms similar to that of the current HLT, including the use of precise momentum measurements. Facilitating tracking in the L1 trigger is an important driver of the design of the Phase-II Tracker. The upgraded L1 “track trigger” will require a new hardware architecture to incorporate the tracking information. While the addition of track information in the L1 trigger provides significant gains in rate reduction with good efficiency, it will nevertheless be necessary to increase the trigger accept rate in order to maintain the required efficiency for all of the important physics channels. This is particularly the case for triggers involving hadrons and photons, for which the sensitivity to PU is higher and the track trigger is somewhat less efficient.

The measurement of processes with small production cross-sections and/or decay branching ratios is a major goal of the HL-LHC physics program. This requires specific upgrades in the forward regions of the detector to maximize the physics acceptance over the largest solid angle possible. To ensure proper trigger performance within the present coverage, the muon system will be completed with new chambers. The new endcap calorimeter configuration offers the opportunity to extend the muon coverage with a tagging station up to  $|\eta| \approx 3$  or more, with significant acceptance gain for multi-muon final states. To mitigate PU effects in jet identification and energy measurement, the tracker will be extended up to  $|\eta| \approx 4$ , thereby also covering the peak production region of jets accompanying Vector Boson Fusion (VBF) and Vector Boson Scattering (VBS) processes, which are among the highest priorities of the physics program. With this extension, measurements of total energy and missing energy will be greatly improved, and b-tagging acceptance will be increased.

As the luminosity integrated over the Phase-II operation period will not be limited by the accelerator performance but by the ability of the detector to sustain high PU, the upgrades of the readout electronics will be designed with some margin to allow efficient data taking up to a PU of 200. This will also provide some flexibility for the luminosity levelling process to use instantaneous luminosity and corresponding beam lifetime information to maximize the useful integrated luminosity in order to obtain the best possible physics performance of the experiment. It is expected that the sustainable luminosity limit will be driven by the performance of sub-detectors that are not going to be replaced for Phase-II. Further simulation studies will include possible alternatives to the luminosity levelling schemes and PU beyond 140 to further optimize designs and to determine the best operating scenarios. Among the studies foreseen, CMS will investigate whether or not precise measurements of the production time of particles, which vary with a rms of  $\sim 200$  ps within a single bunch crossing, would enable a valuable im-



provement in PU mitigation, particularly for the contributions of neutral particles, which are not detected in the tracker.

### 2.2.5 Elements of the Phase-II Upgrades

**Tracker** The Tracker will suffer significant radiation damage by LS3 and must be completely replaced for Phase-II. To maintain adequate track reconstruction performance at the much higher PU levels of the HL-LHC, the granularity of both the outer tracker and the pixel systems will be increased by roughly a factor 4. In the outer tracker, this will be achieved by shortening the lengths of silicon sensor strips relative to those in the current detector, without changing the pitch very significantly. A number of design improvements will lead to a much lighter Outer Tracker providing significantly improved  $p_T$  resolution and a lower rate of  $\gamma$ -conversions compared to the present detector. In addition, the module design will be capable of providing track-stub information to the L1 trigger at 40 MHz for tracks with  $p_T \geq 2 \text{ GeV}$ . This will ensure powerful background rejection at the earliest stage of the event selection. The pixel system will implement smaller pixels and thinner sensors for improved impact parameter resolution and better two-track separation. This will improve b-tagging as well as  $\tau$ -hadronic decay and track reconstruction efficiencies within boosted jets. With up to 10 additional pixel disks in each of the forward regions the system coverage will be extended to close to  $|\eta| = 4$ , to better match the range of coverage of the calorimetry.

**Calorimeter endcaps** The electromagnetic and hadronic endcap calorimeters will also suffer significant radiation damage by LS3, and so must be replaced.

The replacement is called the High Granularity Calorimeter (HGC) and has electromagnetic and hadronic sections with excellent transverse and longitudinal segmentation. It will provide detailed three dimensional images of showers. The electromagnetic section consists of  $\sim 30$  tungsten and copper plates interleaved with silicon sensors as the active material. The sensors have pads of variable sizes of less than  $\sim 1.0 \text{ cm}^2$ . The electromagnetic section has  $25X_0$  and one interaction length ( $\lambda$ ). The hadronic part has a front section of 12 brass and copper plates interleaved with silicon sensors for a depth of  $3.5\lambda$ . This covers the hadronic shower maximum measurement. It is followed by a “backing hadron calorimeter” of similar design to the current HE detector, brass plates interleaved with plastic scintillating tiles read out with a wavelength shifting fiber, to provide an overall depth of  $\sim 10\lambda$  for the full calorimeter. The design of the High Granularity Calorimeter draws upon the ILC/CALICE[15] concepts for 3D measurement of shower topologies.

**Muon endcaps** The muon system in the region  $1.5 \leq |\eta| \leq 2.4$  currently consists of four stations of Cathode Strip Chambers (CSC). It is the only region of the muon detector that lacks redundant coverage despite the fact that it is a challenging region for muons in terms of backgrounds and momentum resolution. To maintain good L1 muon trigger acceptance in this region it is therefore proposed to enhance these four stations with additional chambers that make use of new detector technologies with higher rate capability, along the lines of what was planned in the original design of CMS. The two first stations are in a region where the magnetic field is still reasonably high and so will use Gas Electron Multiplier (GEM) chambers for good position resolution in order to improve momentum resolution for the standalone muon trigger and to improve the matching with tracks in the global muon trigger. The two last stations will use low-resistivity Resistive Plate Chambers (RPC) with lower granularity but good timing resolution to mitigate background effects. In addition, the implementation of a GEM station in the space that becomes free behind the new endcap calorimeters is being proposed in order to increase the coverage for muon detection to  $|\eta| \approx 3$ .

**Beam radiation protection and luminosity measurement** The systems that provide protection against beam background and measurement of the luminosity will require work in several areas to manage the high radiation levels of the HL-LHC. The protection systems will be upgraded with new poly-crystalline diamond sensors that will be read out using the standard LHC Beam Loss Monitor hardware and software and fully integrated into the LHC control system. The Machine Induced Background (MIB) and Luminosity measuring systems in the Pixel volume must also be replaced.

**Trigger** The latency of the present L1 trigger is limited to  $3.4\ \mu\text{s}$  by the tracker readout. For Phase-II operation, it will be increased to  $12.5\ \mu\text{s}$  to provide sufficient time for the hardware track reconstruction and matching of tracks to muons and calorimeter information. This change will require upgrades of the readout electronics in some of the existing sub-detectors that will be kept for Phase-II. A proper design of the front-end electronics for these systems will allow latency limitations to be overcome and at the same time to eliminate L1-trigger rate restrictions. Based on the expected performance of the trigger with track information, the proposed L1-trigger acceptance rate is 500 kHz for beam conditions yielding 140 PU. This will allow CMS to maintain thresholds comparable to those that will be used in a typical Phase-I trigger menu. To retain comparable performance in beam conditions that result in 200 PU, the L1 rate must increase to 750 kHz, and so all detectors will have readout capabilities compatible with this possibility. Studies are underway to optimize scenarios for the trigger menu and to determine if a higher acceptance rate would further improve the exploitation of the higher luminosity for key physics signals. Any further increase of the L1 readout rate would require an increase of the Pixel readout bandwidth.

Specific sub-detector upgrades also required for CMS to meet these trigger requirements are the front-end electronics of the barrel calorimeter; the Muon readout electronics in the CSCs of the inner rings in stations 2 to 4; and the DT readout.

**Data Acquisition and Trigger Control** The Data Acquisition (DAQ) system will be upgraded to implement the increase of bandwidth and computing power that will be required to accommodate the larger event size and L1-trigger rate, and the greater complexity of the reconstruction at high PU. Compared to Phase-I, the bandwidth and the computing power requirements would respectively increase by factors of about 10(15) and 15(30) for operation at PU of 140(200). This is well within the projected network and computing technology capabilities expected at the time of Phase-II. Assuming an online event selection of 1/100 event at the HLT, as is the case in the current system, the subsequent rate of recorded data will increase at PU of 140(200) to 5(7.5) kHz from LHC Run-I levels of roughly a few hundred Hz.

**Software and Computing** Assuming only technology improvements and maintaining existing techniques, the offline software and computing areas would fall short by a factor of 4(12) of the resources needed for the challenging conditions expected in Phase-II at 140(200) pileup. To minimize the computing needs, both at the online and offline levels, a significant R&D program has started as part of the upgrade effort to improve the algorithms and approaches used for data reconstruction, analysis, storage, and access and to adapt the CMS software and computing model to new technologies and resources.

**Experimental Area and Shutdown Considerations** During long shutdowns CMS is highly configurable to allow access to the various sub-systems, as shown in Fig. 2.8, but the access to different areas must often be sequential because of the limited overall size of the experimental cavern. Shutdown planning for Phase-II is still at an early stage, but an initial evaluation of the work sequence and time estimates indicates that the full scope of work can be accomplished in a shutdown of approximately 30 months duration, from end to re-start of beam operations.

In order to gain flexibility in scheduling the work during LS3 while reducing overall costs, consideration is being given to advancing some specific tasks to LS2, if funding is available. Radiation protection and dose to personnel will be a primary concern in planning the upgrades and the shutdown work. This will require development of special shielding, tooling, and work procedures.

### 2.2.6 Upgrade Performance Studies

Full simulations of detector signals using GEANT 4 have been produced in order to develop the CMS scope for Phase-II and to evaluate the performance of the proposed upgrades. It is assumed that the luminous region has a Gaussian shape along the beam (z) axis with a 5 cm RMS. The configurations that have been simulated are the following:

- the Phase-I detector operated at 50 PU ( $10^{34}\text{cm}^{-2}\text{s}^{-1}$ ) without radiation aging; to establish a benchmark for the required performance of the Phase-II upgrades;
- the Phase-I detector operated at 140 PU ( $5 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$ ) with modelling of the effects of radiation damage after integrated luminosities of  $1000\text{fb}^{-1}$ , and in a few cases up to  $3000\text{fb}^{-1}$ , for the outer tracker and hadron and electromagnetic calorimetry in order to identify the key areas to be addressed by the upgrades, and to also determine when the need to install upgrades will become critical; and
- the Phase-II detector operated at 140 PU ( $5 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$ ); to evaluate the performance reach for the new concepts. It is assumed that the performance of the new sub-detectors will not degrade with radiation while an intermediate aging of  $1000\text{fb}^{-1}$  is included for the barrel calorimetry since their active elements will not be replaced.

Since the pixel design is still being developed, the Phase-I configuration has been used for the Phase-II simulations, implementing the new disks in the forward regions but not incorporating expected resolution improvements. Software development for the reconstruction of physics objects (leptons, photons, jets, total hadronic and missing energy), including the sophisticated particle flow techniques for global event reconstruction, is a long-term endeavor. The CMS collaboration has made a large effort to adapt existing software to the new detector geometries and to improve or tune algorithms for the higher PU conditions. This work is still ongoing. Taking these considerations into account, it follows that the performance presented in this document should be considered to be conservative.

It is also important to recognize that the second configuration does not describe the full aging of the detector at  $1000\text{fb}^{-1}$ . In reality, the first two layers of the pixel detector will be too damaged to operate and the forward electromagnetic and hadronic calorimetry will also be non-functional. The outer tracker will also be at the point at  $1000\text{fb}^{-1}$  where it will soon become completely dysfunctional.

Simulations based on DELPHES [16], implementing simplified detector geometries and parameterization of efficiencies and resolutions applied at the generator level, have also been used to produce samples with sufficient statistical precision for studies of the physics backgrounds. The DELPHES parameterizations and results have been tuned and verified by comparison of various DELPHES samples with corresponding samples that have been produced with the full CMS GEANT4-based simulation.

## 2.3 Project Organization and Management

- Overview
- International CMS and CMS Upgrade Project
- US CMS
- Fermilab as host lab
- Other Institutional Partners
- Funding Agencies

## 2.4 Key Performance Parameters

## 2.5 Cost and Schedule

- (including Life Cycle Cost assumptions)

DRAFT

## Chapter 3

# Science - Editor Hill 30 pages

### 3.1 Physics Goals to Physics Requirements

As introduced in Section , the physics program of the HL-LHC is to achieve the following science goals:

1. To perform few percent measurement of all Higgs couplings
2. To find the BSM solution to Hierarchy Problem (e.g. SUSY), or conclude SM is fine-tuned
3. To illuminate particle nature of Dark Matter
4. To use the large dataset to study rare SM processes

In this section, we present the studies carried out by the CMS collaboration that were used to define the science requirements for the the HL-LHC upgrade to the CMS detector that are necessary to achieve these science goals. These science requirement will be expanded on in the next section.

### 3.2 Precision Measurements of the Higgs Couplings

In 2012 the LHC experiments ATLAS and CMS observed a new particle consistent with the properties of a Higgs boson [17–19] with a mass of about 125 GeV. This discovery strongly suggested the validity of the Higgs mechanism as the source of electroweak symmetry breaking (EWSB) and mass generation for fundamental particles. Since the discovery, this Higgs bosons properties, in particular the couplings of the fundamental particles to it, have been measured to a precision of the order of 20%. The large datasets of the HL-LHC are needed to increase the precision of these measurements to a few percent in order to provide a stringent test of the Standard of the Higgs mechanism’s role in EWSB and mass generation in the Standard Model (SM). Studies [20, 21] indicate that this precision is needed in order to discriminate between the SM and a broad class of new physics scenarios.

CMS has reported projections on the expected sensitivity to the Higgs boson couplings [4]. These are based on the measurements that were made in 2013, extrapolated to the  $3000 \text{ fb}^{-1}$  HL-LHC dataset and for a center-of-mass energy of 14 TeV, by scaling the signal and background yields accordingly. The estimated precision on the measurements for modified couplings for a SM-like Higgs boson is shown in Figure 3.1. Two scenarios are used to extrapolate uncertainties, namely scenario 1 where all systematic uncertainties are left unchanged and scenario 2

where the theoretical uncertainties are reduced by a factor of two and all other systematic uncertainties are reduced by the square root of the integrated luminosity. The two scenarios are expected to bracket a realistic extrapolation.

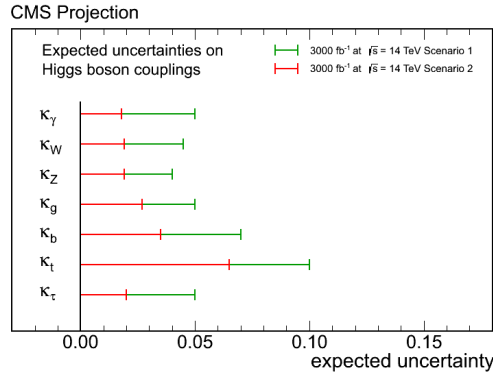


Figure 3.1: Estimated precision on the measurements for modified couplings for a SM-like Higgs boson [4]. The projections assume a center-of-mass energy of 14 TeV and a dataset with integrated luminosity of 3000 fb<sup>-1</sup>. The projections are obtained with two uncertainty scenarios as described in the text.

### 3.2.1 $H \rightarrow ZZ^* \rightarrow 4\ell$ analysis

The Higgs boson decay into two Z bosons, each decaying into two charged leptons, electrons or muons, is the golden channel in the study of the Higgs boson. Electrons and muons can be measured very accurately, with high efficiency, and excellent energy and momentum resolution. The complete final state of the Higgs boson decay can be reconstructed which leads to a signal of high purity, measured as a peak over a smooth background distribution. The four lepton events allow for a detailed CP analysis of the Higgs particle by measuring angular distributions, such as the angle between the ZZ decay planes and the decay angles in these planes, which contain information of the CP properties of the parent boson. The analysis of the full four lepton mass spectrum contains information of the total width of the Higgs boson.

The HL-LHC will produce about sixteen thousand Higgs boson events per experiment in this channel through the gluon-fusion production process and about 1400 through vector boson fusion. Selecting these events with the largest possible acceptance is crucial to the Higgs physics program. Excellent electron/muon triggering efficiency and subsequent reconstruction at low transverse momentum over as large a pseudorapidity coverage as possible are essential. Since four leptons need to be reconstructed in this final state, any single object inefficiency is potentiated.

### 3.2.2 $H \rightarrow \mu\mu$ analysis

The coupling of the Higgs boson to third-generation fermions was established during Run-I of the LHC [22]. The HL-LHC will give unique access to Higgs boson couplings to second-generation fermions. Measurements of the couplings of the Higgs boson to the second generation are more challenging, as a result of their smaller values and hence smaller experimental rates. The most promising channel is the search for the decay of a Higgs boson in two muons with a branching fraction of  $2.2 \times 10^{-4}$  expected in the SM.

In  $H \rightarrow \mu^+\mu^-$  events the kinematics of the Higgs boson can be fully reconstructed. The signal will consist of a small bump over a large di-muon background from Drell-Yan events, hence an



excellent di-muon mass resolution is crucial. This puts constraints on the required performance of the new tracking system.

The measurement of the Higgs boson coupling to muons is expected to improve with the square-root of the improvement in resolution and efficiency. Based on previous projections, an uncertainty in the Higgs boson coupling to muons of about 5 % is expected.

### 3.2.3 $H \rightarrow \tau\tau$ analysis

Projections of the Run-I  $H \rightarrow \tau\tau$  analysis show that the coupling modification of the Higgs boson to tau leptons with respect to the SM expectation can be measured with a precision of 2–5 %. Modifications of the Higgs boson couplings to fermions of this scale or larger are expected from some BSM Higgs models, particularly those predicting multiple Higgs doublets.

The  $H \rightarrow \tau\tau$  measurements rely strongly on the performance of almost all aspects of the CMS detector. The list of objects used in the analysis ranges from electrons, muons, hadronic taus, jets, b-tagged jets, to missing transverse energy. High efficiencies and low misidentification rates are crucial to control challenging backgrounds and to explore the  $H \rightarrow \tau\tau$  decay in full.

To improve the signal to background ratio, events are categorized based on the number of reconstructed jets and the VBF production signature is exploited. Additional jets from pileup collisions dilute the analysis performance unless they are identified and rejected. This can be achieved by matching the charged constituents of reconstructed jets to the primary vertex.

The leptons and pions from tau decays in Higgs boson events as well as VBF tagging jets have a typical  $p_T$  of 20 to 50 GeV. Excellent trigger efficiency at these thresholds are therefore crucial. To achieve this, tracking information at the Level-1 trigger level will be essential. Adding track information at the trigger level will allow dedicated triggers to more efficiently reject misidentified jets, which have substantially different structures from hadronic tau decays.

The L1 trigger including the track trigger allows a single tau trigger with an offline threshold at 88 GeV and a di-tau trigger with an offline threshold at 56 GeV for each object with  $|\eta| < 2.4$ . The equivalent trackless trigger thresholds for the same rate would be 140 GeV and 90 GeV, respectively.

Compared to tau triggers without tracking capability, the track-aware scenario increases the absolute acceptance by 550%. A precision Higgs boson needs tau final states and this would be very difficult without the track trigger upgrade.

### 3.2.4 Higgs Cross Section Measurements

Projections of expected sensitivities for Higgs cross section measurements, fiducial and differential cross sections have been performed by scaling the signal and background yields of the 2016 analysis which used  $12.9 \text{ fb}^{-1}$  of data. These measurements are statistically limited but less dependent on model assumptions and with increased integrated luminosity these measurements will become important, allowing for detailed comparison with SM predictions. For the projections of expected sensitivities, two scenarios for the systematic uncertainties have been considered. In Scenario 1, the theoretical and experimental uncertainties are assumed to remain unchanged with respect to the reference analysis until the end of the LHC programme. Only the statistical uncertainty is scaled with additional integrated luminosity. This is a pessimistic scenario since by the end of the LHC programme, it is expected that the theoretical and experimental uncertainties will be better understood. In Scenario 2, certain assumptions are made as to how the theoretical and experimental uncertainties could be reduced by the time of the HL-LHC. For theoretical uncertainties, improvements to generators, such as moving to



NLO-level predictions, are assumed to be able to reduce the size uncertainties by a factor 2. For experimental uncertainties, it is assumed that refined techniques could allow large improvements to the size of certain systematic uncertainties, they will scale as  $1/\sqrt{\mathcal{L}}$  until they reach a defined lower limit based on the achievable accuracy of the upgraded detector. The lower limit for uncertainties on the integrated luminosity, the lepton identification efficiency and jet energy scale are 1.5%, 1% per lepton, and 1% respectively. Projections to 300 and 3000 fb<sup>-1</sup> are considered, and for the 3000 fb<sup>-1</sup> case the signal and background yields are adjusted to take into account effects of an upgraded detector and higher pileup according to studies in the PhaseII technical proposal [23].

### 3.2.4.1 $H \rightarrow \gamma\gamma$ fiducial cross section ( $\sigma_{\text{fid}}$ ) measurement

The projected relative uncertainties for the  $H \rightarrow \gamma\gamma$  fiducial cross section ( $\sigma_{\text{fid}}$ ) [24], are shown in Fig. 3.2. The fiducial volume is defined on generator-level quantities, and makes the following requirements on photons: the transverse momentum of the lead (sublead) photon,  $p_T^{\text{gen}}(\gamma_{1(2)})$ , should be greater than 1/3 (1/4) of the invariant mass of the diphoton system ( $m_{\gamma\gamma}$ ); the absolute value of the pseudorapidity of both photons ( $|\eta^{\text{gen}}(\gamma_{1,2})|$ ) should be less than 2.5; the isolation of the photons ( $\text{Iso}_{R=0.3}^{\text{gen}}(\gamma_{1,2})$ ), calculated as the sum of the transverse momenta of all stable particles inside a cone of aperture  $R = 0.3$  around the photon, is required to be less than 10 GeV. Projections are given for 300 fb<sup>-1</sup> and 3000 fb<sup>-1</sup> under different scenarios. For the 300 fb<sup>-1</sup> case, the scenarios are S1 and S2, while for the 3000 fb<sup>-1</sup> case, the scenarios are S1+ and S2+, where the “+” indicates that pileup effects have been taken included. The effect of the high pileup conditions of the HL-LHC have been taken into account as degradations to the photon identification efficiency and vertex identification efficiency, following earlier studies [23]. In each case the uncertainty for the inclusive signal strength is also shown split into components: statistical uncertainties (‘stat.’), experimental systematic uncertainties (‘exp.’).

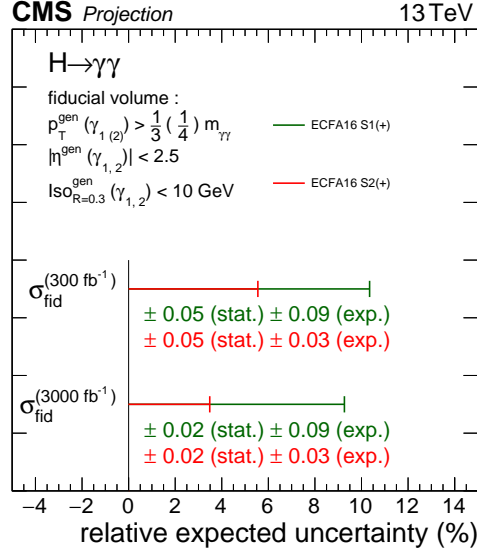
### 3.2.4.2 $H \rightarrow ZZ$ transverse momentum differential cross section measurement

In Fig. 3.3, the projection of the differential cross section measurement as a function of the transverse momentum of the higgs boson is shown. This projection is based on the  $H \rightarrow ZZ \rightarrow 4\ell$  decay channel [25]. The cross section is measured in a fiducial phase space closely matching the experimental acceptance. The projections for 3000 fb<sup>-1</sup> use different lepton efficiencies and misidentification rates to account for the higher pileup at the HL-LHC. In Scenario 2, the experimental uncertainties on the integrated luminosity and the lepton identification efficiency are reduced to 1.5% and 1% per lepton, respectively. The statistical uncertainty of the measurement ranges from 10–29% (4–9%) for 300 (3000) fb<sup>-1</sup> and the high  $p_T$  region is still dominated by statistics at 3000 fb<sup>-1</sup>. The theoretical uncertainty on the differential gluon fusion cross section, which does not affect the measurement, is taken at NLO and shown in magenta.

## 3.2.5 Higgs boson pair production

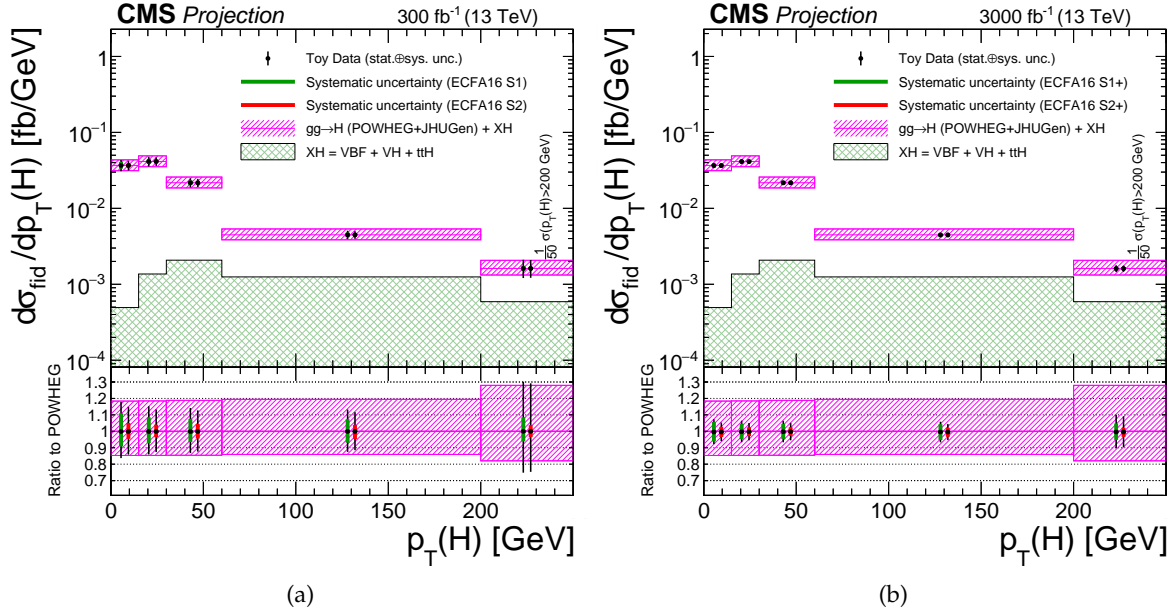
Studies of Higgs boson pair production at the HL-LHC will provide insight on Higgs boson trilinear coupling [26]. This measurement directly probes the Higgs potential. Figure 3.4 shows the dominant Feynman diagrams. Di-Higgs events can be produced via a box diagram and through the Higgs boson self-coupling contribution. The two processes interfere destructively and the cross section is near minimum for the SM. The NNLO pair production cross section of 40.7 fb [27] is about 1000 times smaller than the single Higgs boson production cross section.

Studies have been performed of di-Higgs production and decay into  $bb\gamma\gamma$ ,  $bbWW$ , where the  $W$  boson decays leptonically, and  $bb\tau\tau$ . It is crucial that the Phase-II detector can cope with the challenging environment of HL-LHC, as pileup mitigation, b-tagging, tau-tagging,



(a)

Figure 3.2: The projected relative uncertainties for the  $H \rightarrow \gamma\gamma$  fiducial cross section ( $\sigma_{\text{fid}}$ ) are shown. For the  $300 \text{ fb}^{-1}$  case, the scenarios are S1 and S2, while for the  $3000 \text{ fb}^{-1}$  case, the scenarios are S1+ and S2+. In each case the uncertainty for the inclusive signal strength is also shown split into components: statistical uncertainties ('stat.'), experimental systematic uncertainties ('exp.').



(a)

(b)

Figure 3.3: Projections for the differential cross section measurement of the Higgs boson transverse momentum at  $300 \text{ fb}^{-1}$  (a) and  $3000 \text{ fb}^{-1}$  (b). The last bin represents the integrated cross section for  $p_T(H) > 200 \text{ GeV}$  and is normalized by  $1/50$  for presentation. The theoretical uncertainty on the differential gluon fusion cross section is shown in magenta.

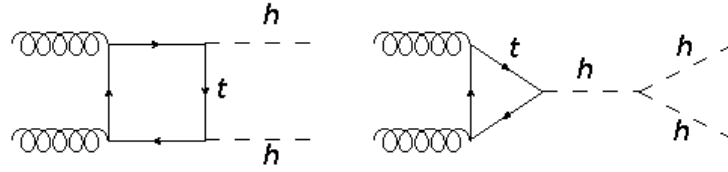


Figure 3.4: Feynman diagrams contributing to gluon fusion Higgs boson pair production.

photon identification efficiencies, and mass resolutions are instrumental to perform these measurements.

### 3.2.5.1 $bb\gamma\gamma$ final state

The signal events of interest contain two high- $p_T$  photons and two high  $p_T$  jets originating from b quarks. Only 320 such events are expected to be produced per experiment at HL-LHC with  $3 \text{ ab}^{-1}$ . The backgrounds can be broadly categorized into resonant backgrounds which contain a Higgs boson decaying to two photons, and non-resonant backgrounds, which do not contain Higgs boson decays. The non-resonant background processes have cross sections that are several orders of magnitude larger than the resonant backgrounds but are suppressed by low rates of mis-tagged jets and mis-identified photons expected for the Phase-II detector.

A two-dimensional maximum likelihood fit of the di-photon,  $M_{\gamma\gamma}$ , and di-b-jet,  $M_{bb}$ , mass distributions is performed to extract the expected signal yield and uncertainty. Figure 3.5 shows the resultant di-photon mass distribution for signal and background processes expected at HL-LHC.

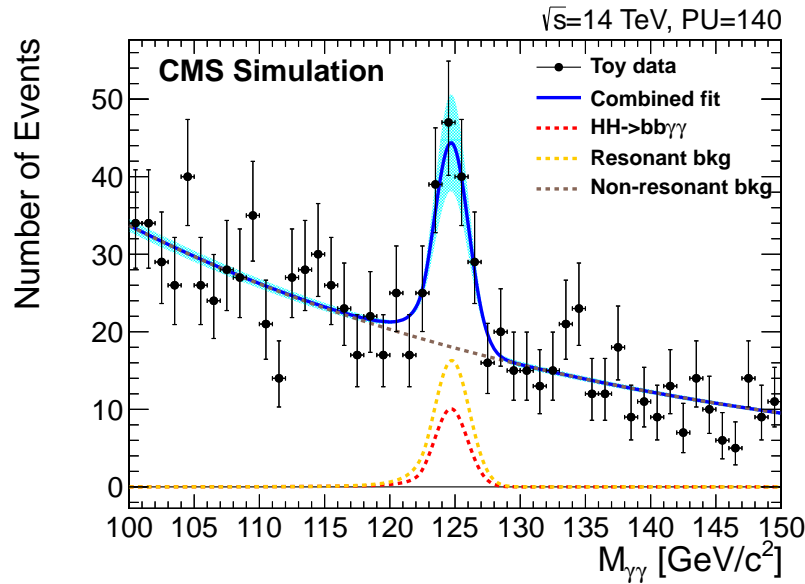


Figure 3.5: Di-photon mass distribution for the estimated signal and background contributions. The data points show the result of a pseudo-experiment.

### 3.2.5.2 $bb\tau\tau$ final state

The  $\tau_\mu\tau_h$ , and  $\tau_h\tau_h$  di-tau final states, where  $\tau_h$  denotes hadronic tau decays, and  $\tau_\mu$  denotes tau decays to muons, are studied. About 9000  $bb\tau\tau$  di-Higgs events per experiment are expected

at HL-LHC with  $3 \text{ ab}^{-1}$ . However, the  $t\bar{t}$  background with fully leptonic decays to taus is overwhelming. Another source of large background is Drell-Yan production of a Z boson decaying into a pair of tau leptons produced in association with jets, where light jets are mis-tagged as b-jets.

The expected significance for di-Higgs boson production above these backgrounds is 0.5, and 0.7 standard deviations, for  $\tau_\mu \tau_h$ , and  $\tau_h \tau_h$  di-tau final states, respectively. For a combination 0.9 standard deviations would be expected. The resulting expected uncertainty in the signal strength is approximately 105%.

The performance of the trigger system is crucial to achieve this result described above, in particular the capability to trigger on charged particles at Level-1. For example, the  $\tau_h \tau_h$  final state, the di-tau trigger has an offline threshold of 56 GeV on both tau legs, and single tau trigger threshold is 88 GeV for the Level-1 sample menu described in this document. These thresholds would be significantly higher without the track trigger, 95 GeV on both tau legs for di-tau trigger and 138 GeV for single tau trigger, which would reduce yields by about a factor of two.

### 3.2.5.3 $bbWW$ final state

About 1500 fully leptonic signal events per experiment are expected at the HL-LHC, where the leptons are either muons or electrons. The dominant background process is dileptonic  $t\bar{t}$ .

After basic kinematic selections are applied, a neural network (NN) discriminator is trained to further reduce the background. The signal selection is obtained by applying a threshold on the NN discriminator leading to 3875 background events and 37.1 signal events. The results suggest a promising contribution of this final state when combined with the other final states at the HL-LHC.

### 3.2.5.4 DiHiggs conclusions

The measurement of Higgs boson pair production at the HL-LHC directly probes the Higgs boson trilinear coupling. The studies performed conclude that an observation of the process is possible in a combination of multiple channels and using results from ATLAS and CMS. Combining the  $bb\gamma\gamma$  and  $bb\tau\tau$  final states, the expected significance for di-Higgs boson production is 1.9 standard deviations.

## 3.3 Searches for SUSY BSM, including DM

The search for BSM physics is a major goal of the LHC physics program. The range of BSM scenarios is broad, and both high-luminosity data samples and the full set of CMS detector capabilities will be needed to provide sensitivity to the signatures, cross sections, and decay branching fractions of interest. To narrow the scope of the discussion, we focus on the search for supersymmetry (SUSY) as the BSM physics in this section. If evidence for one or more new particles is discovered, an extensive program of measurements will be required to determine whether they are indeed SUSY partners of SM particles and to address even more challenging issues, such as discerning the mechanism of SUSY breaking. Because of the complexity of the SUSY spectrum and the associated decay processes, such a program would likely extend for many years, as was the case for studies of the standard-model hadron spectrum. This section presents results from a set of studies [28, 29] that address key questions on the physics potential of the anticipated SUSY program with the  $3000 \text{ fb}^{-1}$  HL-LHC dataset.

The motivations for supersymmetry are strong, in spite of the absence of specific evidence for

this framework. First, SUSY provides a candidate particle, the lightest supersymmetric particle (LSP), that may account for all or part of the astrophysical dark matter. For example, in the minimal supersymmetric standard model (MSSM), the lightest neutralino, designated  $\tilde{\chi}_1^0$ , is a superposition of the spin-1/2 superpartners of the neutral gauge and Higgs bosons, and it can in principle satisfy the constraints from indirect dark matter observations. Second, if a SUSY particle spectrum is present, the three running gauge couplings of the standard model interactions can converge at a common high energy, a requirement for gauge unification. Finally, the discovery of a Higgs boson by the CMS and ATLAS experiments has given new urgency to the gauge hierarchy problem. Assuming that the Higgs boson is a fundamental scalar particle, its mass is extremely sensitive to short distance quantum corrections. Without some kind of new physics that compensates for these effects, the Higgs mass would be pulled up to the Planck scale, barring a coincidental near-perfect cancellation of the bare Higgs mass parameter and the enormous shift induced by quantum corrections. A broad class of SUSY scenarios, known as natural models [30], can stabilize the Higgs boson mass through additional contributions involving diagrams with the scalar superpartner,  $\tilde{t}$ , of the top quark (stop), as well as other SUSY particles.

With the discovery of a Higgs boson, the gauge hierarchy problem is no longer a hypothetical issue, and many searches in the current and future SUSY program are guided by expectations for natural SUSY models. In such models, the top squark, bottom squark, gluino, and higgsino are constrained to be relatively light, while the other superpartner masses are essentially unconstrained. Although the parameter space of natural SUSY models is only loosely defined, the high energy and high integrated luminosity of the HL-LHC program are generally regarded as sufficient to probe most of the relevant model space.

While natural SUSY models are perhaps the best motivated of all scenarios, they are by no means the only models of interest. The studies presented here consider several possibilities to provide a sense of the challenges posed by a range of models.

This discussion of SUSY searches is divided into three sections. Section 3.3.0.1 briefly compares full-spectrum models and simplified models, which provide complementary approaches for interpreting SUSY searches. Section 3.3.1 presents nine separate analysis strategies. These are applied to five full-spectrum benchmark SUSY models, which include three natural SUSY scenarios, as well as stau and stop coannihilation scenarios. The features of these models, and the selection requirements of the nine analyses are briefly described. Section 3.3.2 presents the results obtained from the nine analyses. We consider not only the discovery sensitivity, but also how, in the event of a discovery, the pattern of signals and the associated kinematic distributions can provide many clues to understanding the nature of the underlying particle spectrum. From these studies, it is clear that the full HL-LHC data sample will provide critical information, even if discoveries are made much earlier. Section 3.3.3 summarizes the main results and conclusions of these studies.

### 3.3.0.1 SUSY models used in the interpretation of searches

This section compares full-spectrum SUSY models, which are used in the studies presented in later sections, with simplified-models, which have been used in many of the interpretations of CMS Run 1 data. Figure 3.6 shows the mass spectra for two of the SUSY models, NM3 and STC. The left-most column shows the particles that make up the Higgs sector, with the neutral, CP-even Higgs boson fixed at the observed mass,  $m_H = 125$  GeV. (All five models share this feature, but other aspects of their Higgs sectors can vary from model to model.) SUSY particles with electroweak interactions (only) are shown in the two middle columns. These



are the scalar leptons and scalar neutrinos (sleptons and sneutrinos) and an electroweak sector consisting of the charginos and neutralinos, which are spin-1/2 fermions. The SUSY partners in the electroweak sector, sometimes referred to as ewkinos, are superpositions of higgsinos and gauginos of the same charge. The far-right-hand column shows the strongly interacting sector, with the gluino and the squarks. Because quarks are spin-1/2 objects, they each have two SUSY partners, corresponding to the  $L$ - and  $R$ -handed chiral projections, which have different gauge quantum numbers. Thus,  $\tilde{q}_L$  and  $\tilde{q}_R$  are distinct particles. Mixing is expected to be significant in the third generation of squarks, leading to mass eigenstates designated  $\tilde{t}_1$ ,  $\tilde{t}_2$ ,  $\tilde{b}_1$ , and  $\tilde{b}_2$ .

As we have noted, natural SUSY models are characterized by a small set of relatively light particles. This set consists of  $\tilde{t}_1$ ,  $\tilde{t}_2$ ,  $\tilde{b}_1$  (which in this case is  $\tilde{b}_L$ ), the gluino, and the higgsinos. The masses of the first- and second-generation squarks are not constrained to be light; the sleptons and the electroweak sector, apart from the higgsinos, are also not constrained. While the natural SUSY framework provides important experimental guidance, we will see that the phenomenology can vary substantially across these models.

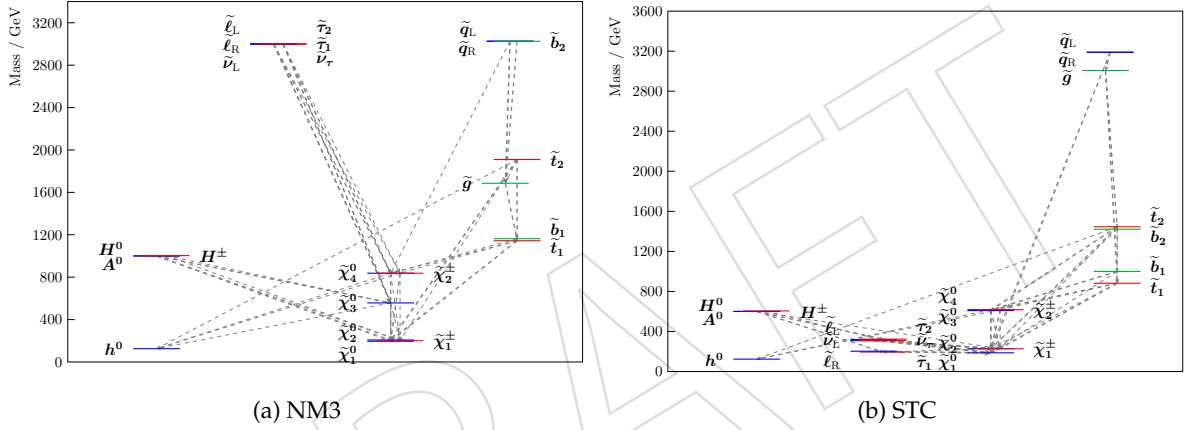


Figure 3.6: Examples of SUSY full-spectrum models: (a) the natural SUSY model NM3 and (b) the stau coannihilation model STC, which are among the five full-spectrum scenarios used in the studies presented here. In NM3, the masses of the  $\tilde{g}$ ,  $\tilde{t}_1$ ,  $\tilde{t}_2$ , and  $\tilde{b}_1$  are all below 2 TeV. The  $\tilde{\chi}_1^0$  is higgsino-like. In the STC model, the gluino is much heavier than the top squarks, and the slepton sector is light, with the  $\tilde{\tau}$  nearly degenerate with the  $\tilde{\chi}_1^0$ . The lines between different states indicate transitions with branching fractions greater than 5%.

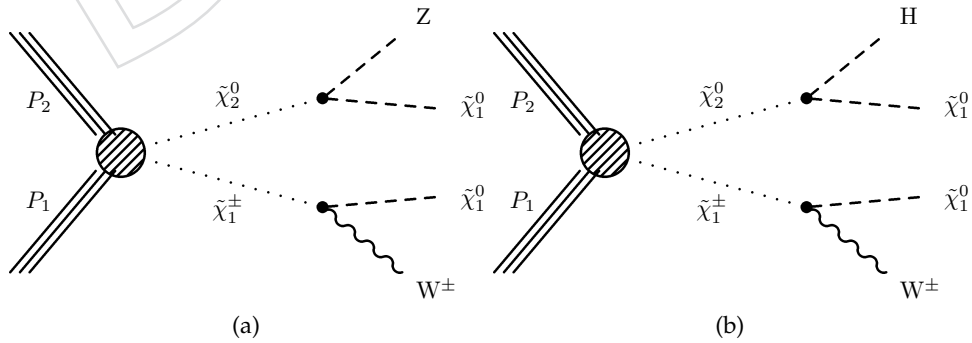


Figure 3.7: Simplified SUSY models for chargino-neutralino production leading to (a)  $W^\pm Z + E_T^{\text{miss}}$  and (b)  $W^\pm H + E_T^{\text{miss}}$  final states.



The interpretation of the results of a SUSY search can be performed either with full-spectrum models, such as the NM3 and STC models shown in Fig. 3.6, or with simplified model spectra (SMS) [31, 32]. Simplified models limit the number of relevant SUSY particles for a particular signature to a minimal number, often just two. For example, a simplified model with just a single top squark,  $\tilde{t}$ , and a lighter neutralino,  $\tilde{\chi}_1^0$ , is sufficient to describe the scenario of top squark pair production with  $\tilde{t}_1 \rightarrow t\tilde{\chi}_1^0$ . Figure 3.7 shows production and decay diagrams associated with two simplified models relevant to the production of a chargino-neutralino pair ( $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ ) pair, with different  $\tilde{\chi}_2^0$  decays, (a)  $\tilde{\chi}_2^0 \rightarrow Z\tilde{\chi}_1^0$  and (b)  $\tilde{\chi}_2^0 \rightarrow H\tilde{\chi}_1^0$ .

Simplified models have been extremely important in helping to provide a systematic and general structure to guide the SUSY search program. In addition, the small number of mass parameters needed to describe such models greatly facilitates scans over the relevant range of the model space. It is important, however, to be aware of the nature of the simplifying assumptions and the limitations that they introduce. For example, limits on the SUSY particle masses could be misleadingly high, because the calculation of these limits requires one to explicitly assume some values for the relevant branching fractions (usually, but not always, taken to be 100%). Interpretations based on both approaches, simplified-spectrum models and full-spectrum models, are discussed here. In the case of conclusions based on simplified models, we have tried to note the key assumptions.

### 3.3.1 Overview of SUSY search strategies and full-spectrum models

The searches presented here span a range of final states ranging from a single high- $p_T$  jet recoiling against  $E_T^{\text{miss}}$  to signatures with three leptons and multiple b-tagged jets. The full SUSY search program involves many more signatures, but these nine studies allow us to illustrate many of the key issues. The details of these searches, including a full description of the selection criteria and tables showing the reduction of backgrounds as the selection requirements are applied, are presented in a separate document [29].

The studies described here were performed using simulated event samples based on two detector configurations. The  $300 \text{ fb}^{-1}$  samples were generated with a simulation of the Phase I detector with 50 pileup interactions [33–35], while for the HL-LHC, a Phase II baseline detector configuration with a pileup of 140 were used.

Monte Carlo simulations samples were produced with these configurations based on GEANT detector simulation, from which the object efficiencies and resolutions were determined and implemented in the DELPHES 3.0.10 fast simulation program [36], which was used for both signal and background production. DELPHES is able to include pileup interactions from inelastic proton-proton interactions simulated with PYTHIA6 [37]. Using MADGRAPH5 [38], samples of 10 to 100 million events per background process were produced. The events include up to four extra partons from initial- and final-state radiation, matched to PYTHIA6 for fragmentation and hadronization. The background cross sections were normalized to next-to-leading-order (NLO) calculations, which is based on the work in preparation for the Snowmass summer study 2013 and discussed in more detail in Refs. [39–41].

The systematic uncertainties assigned in these studies are based partly on those achieved in current 8 TeV analyses. However, the uncertainties are adjusted to reflect their dominant contributions and how these are expected to evolve as the selection procedures are adjusted for higher energy running and higher integrated luminosities. A fuller discussion of these assumptions is presented in Ref. [29].

All of the models are constructed such that their calculated dark matter abundances are at or

below the observed value from WMAP. Two of the models, STC and STOC, involve dark matter coannihilation scenarios in which a second SUSY particle is nearly degenerate in mass with the  $\tilde{\chi}_1^0$ , which helps to keep the dark matter abundance sufficiently low.

### 3.3.1.1 The natural SUSY models (NM1, NM2, NM3)

The natural SUSY models NM1, NM2, and NM3 typify some of the most important scenarios for future searches. The strongly interacting sectors of NM1 and NM2 are nearly identical to that of NM3 (shown in Figure 3.6). In each model,  $m(\tilde{g}) \approx 1.7$  TeV,  $m(\tilde{t}_1) \approx 1.1$  TeV,  $m(\tilde{t}_2) \approx 1.9$  TeV, and  $m(\tilde{b}_1) \approx 1.2$  TeV, whereas  $\tilde{b}_2$  and the first and second generation squarks are much heavier, with masses around 3 TeV.

In each of these models, the gluino pair-production cross section at  $\sqrt{s} = 14$  TeV is  $\sigma(\tilde{g}\tilde{g}) = 5.4$  fb, the largest of any strong process. The cross sections for direct pair production of  $\tilde{t}_1\tilde{t}_1^*$  pairs are in the range  $\sigma(\tilde{t}_1\tilde{t}_1^*) \approx 3 - 4$  fb, while  $\sigma(\tilde{b}_1\tilde{b}_1^*) \approx 3$  fb. Gluino-squark pair production is significant,  $\sigma(\tilde{g}\tilde{q}) \approx 2$  fb in spite of the fact that the first and second generation quarks have large masses,  $m(\tilde{q}) \approx 3$  TeV.

Because gluinos decay strongly and the daughter squarks have similar masses across these models, the gluino branching fractions are also nearly identical in NM1, NM2, and NM3. The decays  $\tilde{g} \rightarrow \tilde{t}_1\bar{t}$  and  $\tilde{g} \rightarrow \tilde{b}_1\bar{b}$  dominate, with branching fractions of 60% and 40%, respectively (conjugate modes are implied). The subsequent decays of  $\tilde{t}_1$  and  $\tilde{b}_1$  lead to final states with four b-quarks, highlighting the importance of b-jet tagging. (The STOC model is constructed to provide an interesting alternative, in which  $\tilde{t}_1 \rightarrow c\tilde{\chi}_1^0$  with nearly 100% branching fraction.)

The electroweak and leptonic SUSY sectors of the natural models differ, however, with dramatic consequences for the  $\tilde{t}_1$  and  $\tilde{b}_1$  decay branching fractions, as well as for the decay patterns of colorless particles. These effects arise from different patterns of masses in the electroweak and lepton sectors, as well as from differences in the gauge and higgs content of the neutralinos and charginos. For example, the decay  $\tilde{t}_1 \rightarrow t\tilde{\chi}_1^0$  almost never occurs in NM1 and NM2, but this mode has a 40% branching fraction in NM3. These differences can be traced to the electroweak sector. For example, in NM1 the LSP has mass  $m(\tilde{\chi}_1^0) \approx 420$  GeV and is Higgsino-like, but in NM2 and NM3, it has mass  $m(\tilde{\chi}_1^0) \approx 200$  GeV and is Bino-like.

### 3.3.1.2 Stau coannihilation model (STC)

In the stau coannihilation model [42], all of the sleptons and sneutrinos are light, and the  $\tilde{\tau}_1$  and  $\tilde{\chi}_1^0$  masses are nearly degenerate, with  $m(\tilde{\tau}_1) = 194$  GeV and  $m(\tilde{\chi}_1^0) = 187$  GeV. The mass degeneracy of the  $\tilde{\tau}_1$  and the  $\tilde{\chi}_1^0$  allows efficient co-annihilation of dark matter to lower the predicted relic density to its observed value.

The gluino is heavy ( $m(\tilde{g}) \approx 3$  TeV), suppressing its production; this is the only benchmark model in which the gluino production cross section is effectively negligible. However, the  $\tilde{t}_1$  and  $\tilde{b}_1$  are relatively light, with  $m(\tilde{t}_1) \approx 880$  GeV and  $m(\tilde{b}_1) \approx 1$  TeV. These masses lead to significant direct pair-production cross sections:  $\sigma(\tilde{t}_1\tilde{t}_1^*) \approx 19$  fb and  $\sigma(\tilde{b}_1\tilde{b}_1^*) \approx 8.3$  fb. (Antiparticles of SUSY partners are denoted here with an asterisk, which does not refer to an off-shell particle.) The decay  $\tilde{b}_1 \rightarrow b\tilde{\chi}_1^0$  has a large branching fraction ( $\approx 70\%$ ) and provides a key signature in an all-hadronic search. As a consequence of  $\tilde{t}_1$  decays to the low-mass electroweak sector, the single-lepton  $\tilde{t}_1$  search and the triplepton +b-jet search both provide good sensitivity.

The STC model lies in the parameter space that yielded the highest likelihood in fits to all pre-LHC experimental data within the constrained MSSM (cMSSM) [42]. These fits preferred

scenarios with a small mass difference, about 10 GeV, between the  $\tilde{\tau}$ -NLSP and the  $\tilde{\chi}_1^0$  as LSP. Within the context of the cMSSM, this region is ruled out by LHC searches based on the strongly interacting sector, which in the cMSSM is coupled to the electroweak sector by constraints arising from GUT-scale mass unification. However, without these cMSSM-specific constraints, the part of the spectrum that is closely coupled in the fit to electroweak and flavor precision observables and to dark matter is not in conflict with existing LHC results.

### 3.3.1.3 Stop coannihilation model (STOC)

The stop coannihilation [43, 44] model STOC is also formulated in the cMSSM parameter space. In this model,  $m(\tilde{t}_1) \approx 400$  GeV is very low, and the top squark is nearly degenerate with the  $\tilde{\chi}_1^0$ , which is bino-like. As a consequence, the direct top squark pair production cross section is enormous,  $\sigma(\tilde{t}\tilde{t}^*) \approx 2.1$  pb. The top squark decays are effectively invisible, however, because they proceed via the loop process  $\tilde{t} \rightarrow c\tilde{\chi}_1^0$ , in which the daughter charm jet is extremely soft due to the small mass splitting between  $\tilde{t}_1$  and  $\tilde{\chi}_1^0$ . Nevertheless, if the  $\tilde{t}_1$ -pair system is boosted against a hard jet from initial-state radiation, the process is experimentally accessible in the single jet +  $E_T^{\text{miss}}$  signature, as in monojet searches [45].

In the STOC model, the gluino is the second lightest strongly interacting particle, with  $m(\tilde{g}) \approx 2.1$  TeV. The gluino pair-production cross section is  $\sigma(\tilde{g}\tilde{g}) \approx 0.5$  fb, about one-tenth that in NM1–NM3. The squark-gluino cross section is comparable,  $\sigma(\tilde{q}\tilde{g}) \approx 0.3$  fb, and leads to a significant event yield in the all-hadronic  $H_T$ – $H_T^{\text{miss}}$  search.

Although the gluino pair-production cross section is large enough to be experimentally visible, the decay signatures are quite distinct from those in the natural SUSY models. This may be surprising, given that the gluino decays with 50% branching fractions each to  $\tilde{t}_1\bar{t}$  and  $\tilde{b}_1\bar{b}$ , which are comparable to those in the natural models. As we have seen, however,  $\tilde{t}_1$  decays are essentially invisible in STOC, so gluino pair production with  $\tilde{g} \rightarrow \tilde{t}\bar{t}$  simply yields two top quarks plus large  $E_T^{\text{miss}}$ . In summary, for STOC, the discoveries of signals in the monojet and  $\tilde{t}\bar{t} + E_T^{\text{miss}}$  searches, combined with the absence of new physics in the  $4b + E_T^{\text{miss}}$  signature, could suggest a scenario of this type.

## 3.3.2 Results from sensitivity studies

### 3.3.2.1 Results for the SUSY models NM1, NM2, and NM3

There are both strong similarities, as well as some distinctive features, in the pattern of expected search results for NM1, NM2, and NM3. As discussed in Section 3.3.1.1, these models have nearly identical strongly interacting sectors, including relatively light  $\tilde{g}$ ,  $\tilde{t}_1$ ,  $\tilde{t}_2$ , and  $\tilde{b}_1$ . However, these models differ significantly in the properties of their electroweak and leptonic sectors.

The all-hadronic  $M_{T2}$  analysis with a requirement of  $\geq 3b$ -tags has strong discovery potential for these models. The  $M_{T2}$  variable (also referred to as stransverse mass) is described in the references [46, 47]. The signals arise primarily from gluino pair production, with  $\tilde{g} \rightarrow \tilde{t}\bar{t}$  and  $\tilde{g} \rightarrow \tilde{b}\bar{b}$ , as discussed in Sec. 3.3.1.1. Figure 3.8(a) shows the distribution of the  $M_{T2}$  for each of the three natural SUSY models and for the SM backgrounds. The dominant SM background arises from  $\tilde{t}\bar{t}$  production, but these events are concentrated at lower values of  $M_{T2}$ . Significant signals are observed in the higher  $M_{T2}$  regions for each of the scenarios, although the yields vary substantially over the natural models. The slopes of the distributions are also different and reflect the masses and decay patterns.

The resulting sensitivity for NM2 and NM3 already exceeds  $5\sigma$  at  $300 \text{ fb}^{-1}$ , while that for NM1 is lower, requiring roughly  $2000 \text{ fb}^{-1}$  to reach this significance. The lower sensitivity to NM1

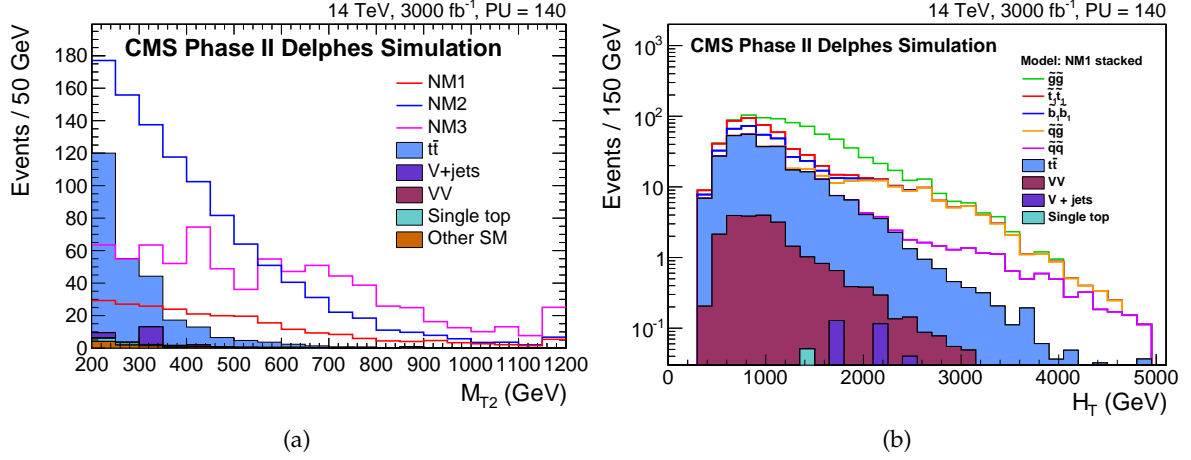


Figure 3.8: Distributions from two searches with sensitivity to the NM1, NM2, and NM3 models. (a) All-hadronic search with  $M_{T2}$ : the distribution of  $M_{T2}$  after the full selection (including  $\geq 3$  b-tags) except the  $M_{T2}$  requirement itself. (b) Single-lepton search: distribution of  $H_T$  after the selection requirements are applied to all other variables. The contributions of the SM backgrounds are shown as stacked histograms in both (a) and (b), as they are elsewhere in this paper. In (a), the SUSY signal contributions from the different models are shown overlaid, but in (b) they are shown stacked because the histograms represent different processes within a single SUSY scenario, NM1.

in the all-hadronic search arises from the properties of its electroweak sector, which produces more leptons. Because events with observed leptons are vetoed in an all-hadronic analysis, the average efficiency for events in this model is lower than that for NM2 and NM3. This feature illustrates the complementarity of different search signatures. The mass splittings in a given model also have strong effects on the sensitivity. In NM3, the lower  $\tilde{\chi}_2^0$  and  $\tilde{\chi}_1^\pm$  masses lead to a harder  $M_{T2}$  distribution. This distinctive feature boosts the discovery sensitivity for NM3.

The single-lepton search also has strong sensitivity to each of the natural SUSY models. Figure 3.8(b) shows the distribution of  $H_T$ , the scalar sum of the jet  $p_T$  values, for the single-lepton search. Rather than showing the signal distributions for each of the natural SUSY models, only NM1 is shown, but the contributions from the various SUSY production processes within NM1 are displayed. For most of the  $H_T$  region, up to around 2 TeV, the signal yield is dominated by gluino-pair production, but at high  $H_T$ , above 2.5 TeV, gluino-squark production becomes prominent. The flavor of the produced squark is inherited from that of the incoming quark from the proton, so  $\tilde{u}$  and  $\tilde{d}$  dominate. It is remarkable that, with a 3000 fb<sup>-1</sup> sample, this contribution is observable even though the masses of the  $\tilde{u}$ - and  $\tilde{d}$ -squarks are 3 TeV. In fact, the presence of contributions from multiple SUSY decay chains often complicates the extraction of a signal from a particular processes of interest. This search is nominally designed for direct  $\tilde{t}\bar{t}$  pair production, but Fig. 3.8(b) shows that other SUSY processes would dominate the observed yields in this scenario.

The trilepton + b-tag search also has good sensitivity to NM1, NM2, and NM3. The events are separated into two bins with b-tag multiplicity of either 2–3 or  $\geq 4$ . The higher multiplicity bin is well suited to searching for gluino pair production in which each gluino decays into either a  $\tilde{t}\bar{t}$  or  $\tilde{b}\bar{b}$  pair. Figure 3.9(a) shows the distribution of  $E_T^{\text{miss}}$  in the trilepton +  $\geq 4b$  event sample, which is well-suited for NM1 and NM2. In NM3, the b-jets are softer due to smaller mass splittings, and the 2–3 b-jet multiplicity bin provides better sensitivity.

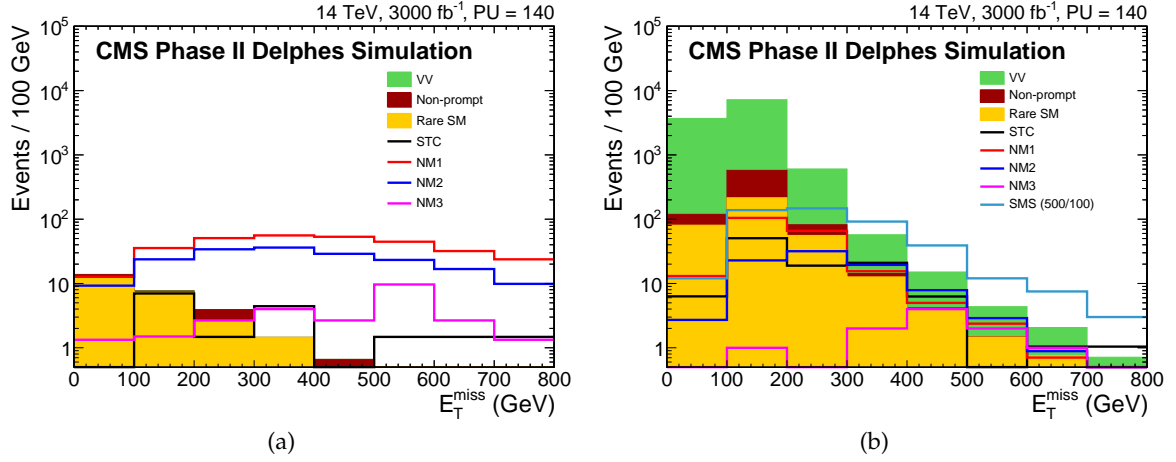


Figure 3.9: Trilepton searches with and without b-jet tagging. (a) Distribution of  $E_T^{\text{miss}}$  for the trileptons + b-jets search in the bin with  $\geq 4$  b-tagged jets. (b) Distribution of  $E_T^{\text{miss}}$  in the trilepton search for  $\tilde{\chi}_1^\pm \tilde{\chi}_1^0$ , with  $\tilde{\chi}_1^\pm \rightarrow W^\pm \tilde{\chi}_1^0$  and  $\tilde{\chi}_2^0 \rightarrow Z \tilde{\chi}_1^0$ . In this search, no b-tagging requirements are used. The selected events satisfy  $200 < M_T < 400$  GeV and have an  $\ell^+ \ell^-$  pair that reconstructs to the Z-boson mass.

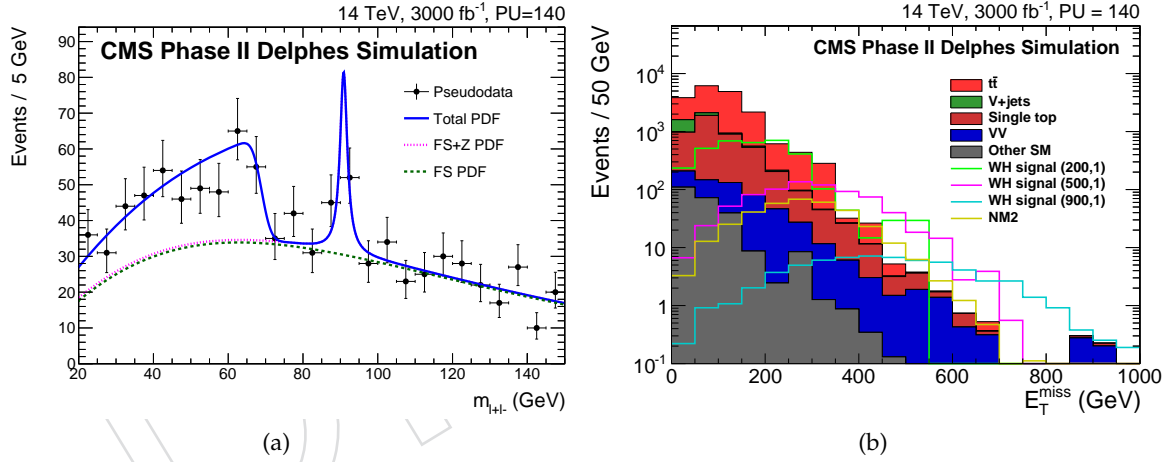


Figure 3.10: (a) Distribution of the dilepton invariant mass  $m_{\ell^+ \ell^-}$  in NM1, showing the distinctive kinematic edge associated with the decay chain  $\tilde{\chi}^\pm \rightarrow \tilde{\ell}_L^\pm \ell^\mp$ ;  $\tilde{\ell}_L^\pm \rightarrow \ell^\pm \tilde{\chi}_1^0$ . (b) Search for  $\tilde{\chi}_2^0 \tilde{\chi}_1^\pm \rightarrow WH + E_T^{\text{miss}}$ : distribution of  $E_T^{\text{miss}}$ .

Figure 3.9(b) shows the distribution of  $E_T^{\text{miss}}$  from a trilepton search in which events with b-jets are vetoed. This analysis is designed for processes such as  $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ , with  $\tilde{\chi}_1^\pm \rightarrow W^\pm \tilde{\chi}_1^0$  and  $\tilde{\chi}_2^0 \rightarrow Z \tilde{\chi}_1^0$ , where the signature consists of leptons and  $E_T^{\text{miss}}$  only. For generality, separate signal regions are defined for events with  $\ell^+ \ell^-$  pairs reconstructed on and off the Z-boson mass peak. Figure ?? shows that the trilepton analysis without b-jet tagging provides  $\geq 3\sigma$  sensitivity for NM1, NM2, and STC at  $3000 \text{ fb}^{-1}$ .

Figure 3.10(a) shows the distribution of  $m_{\ell^+ \ell^-}$  in the dilepton-edge analysis of NM1 for the signal sample (opposite-sign, same-flavor leptons), where a prominent triangle-shaped signal excess is evident, with its distinctive edge at the upper end. The location of the edge is determined by the  $\tilde{\chi}_{1,2}^0$  and  $\tilde{\ell}_L$  masses through the relation  $m_{\text{edge}} = \sqrt{(m_{\tilde{\chi}_2^0}^2 - m_{\tilde{\ell}_L}^2)(m_{\tilde{\ell}_L}^2 - m_{\tilde{\chi}_1^0}^2)}/m_{\tilde{\ell}_L}$ .



Most of the background can be estimated from an  $e^\pm\mu^\mp$  control sample. Sensitivity to this critical feature of NM1 requires the full high-luminosity LHC data sample.

The NM2 model produces distinctive signals that arise from  $\tilde{\chi}_1^\pm\tilde{\chi}_2^0$  production. In this model,  $B(\tilde{\chi}_1^\pm \rightarrow W^\pm\tilde{\chi}_1^0) \approx 100\%$ ,  $B(\tilde{\chi}_2^0 \rightarrow H\tilde{\chi}_1^0) \approx 90\%$ , and  $B(\tilde{\chi}_2^0 \rightarrow Z\tilde{\chi}_1^0) \approx 10\%$ , leading to low-multiplicity signatures with leptons,  $E_T^{\text{miss}}$ , and, in the case of  $H \rightarrow b\bar{b}$ , a pair of  $b$  jets with an invariant mass  $m(b\bar{b}) \approx m_H$ . The origin of these signals is the large splitting between  $m(\tilde{\chi}_2^0) \approx m(\tilde{\chi}_1^\pm) \approx 530 \text{ GeV}$  and  $m(\tilde{\chi}_1^0) \approx 200 \text{ GeV}$ , allowing  $\tilde{\chi}_1^\pm \rightarrow W^\pm\tilde{\chi}_1^0$  and  $\tilde{\chi}_1^0 \rightarrow H\tilde{\chi}_1^0$ . Figure 3.10(b) shows the distribution of  $E_T^{\text{miss}}$  in the search using  $W^\pm \rightarrow \ell^\pm\nu$  and  $H \rightarrow b\bar{b}$ . With  $3000 \text{ fb}^{-1}$ , this search is expected to observe a signal for NM2 with 3–5 $\sigma$  significance, as shown in Figure ??.

For processes such as  $\tilde{\chi}_2^0\tilde{\chi}_1^\pm$ , where there are plausible scenarios in which only a small number of particles play a role, it is useful to interpret the results more generically using a simplified model spectrum. The relevant simplified model processes have already been shown in Fig. 3.7(b). Besides the SM background and the signal for NM2, Fig. 3.10(b) shows the signal distributions for three simplified model scenarios. The models involve three SUSY particles but are characterized by only two parameters:  $m(\tilde{\chi}_2^0) = m(\tilde{\chi}_1^\pm)$  and  $m(\tilde{\chi}_1^0)$ . The production cross section decreases with increasing  $m(\tilde{\chi}_2^0) = m(\tilde{\chi}_1^\pm)$ , but for  $m(\tilde{\chi}_1^0)$  fixed essentially at zero, the extent of the high  $E_T^{\text{miss}}$  tail increases. Because there are only two independent model parameters, one can scan over the parameter space and present the results as a region in the plane of  $m(\tilde{\chi}_1^0)$  vs.  $m(\tilde{\chi}_2^0) = m(\tilde{\chi}_1^\pm)$ . Figure 3.11 shows the boundaries of the 5 $\sigma$  discovery regions for data samples of  $300 \text{ fb}^{-1}$  and  $3000 \text{ fb}^{-1}$  under three different assumptions:  $\mathcal{B}(\tilde{\chi}_2^0 \rightarrow Z\tilde{\chi}_1^0) = 100\%$ ,  $\mathcal{B}(\tilde{\chi}_2^0 \rightarrow H\tilde{\chi}_1^0) = 100\%$ , and 50% for each branching fraction. (These scenarios do not span the possible space in a full-spectrum model scenario. Both the  $\tilde{\chi}_1^\pm$  and the  $\tilde{\chi}_2^0$  can in principle decay via additional modes.) The exclusion region from the 2012 analysis is also shown. The new data will greatly expand the discovery region for this key signal process.

### 3.3.2.2 Results for the stau coannihilation model STC

As discussed in Sec. 3.3.1.2, the stau coannihilation model is characterized by a light slepton sector, with the  $\tilde{\tau}_1$  degenerate in mass with the  $\tilde{\chi}_1^0$ ; a relatively light third-generation squark sector; and a heavy, essentially decoupled gluino. As a consequence, the STC model has a very different profile of strong-production cross sections from the natural models, where gluino pair production plays a major role. In STC, the dominant strong-interaction processes are  $\tilde{t}_1\tilde{t}_1^*$  and  $\tilde{b}_1\tilde{b}_1^*$  pair production, with cross sections of  $19 \text{ fb}^{-1}$  and  $8.3 \text{ fb}^{-1}$ , respectively. Figure ?? shows that the all-hadronic search with  $H_T$  and  $H_T^{\text{miss}}$ , the all-hadronic  $\tilde{b}_1$  search, the single-lepton search, the trilepton search, and the trileptons +  $b$ -tag search all provide sensitivity to this scenario.

Figure 3.12 shows distributions of the boost-corrected contranverse mass [48, 49]  $M_{CT}$  from the the all-hadronic search for  $\tilde{b}_1\tilde{b}_1^*$  pair production, which focuses on the signature with 2 $b$ -tagged jets +  $E_T^{\text{miss}}$ . This variable is designed for the analysis of events in which two heavy particles decay into a jet +  $E_T^{\text{miss}}$ . The  $M_{CT}$  distribution for signal events has an endpoint at the value  $M_{CT}^{\text{max}} \approx (m^2(\tilde{b}_1) - m^2(\tilde{\chi}_1^0))/m(\tilde{b}_1)$  [48, 49]. Figure 3.12(a) shows the  $M_{CT}$  distribution for  $M_T > 750 \text{ GeV}$ , while Fig. 3.12(b) shows the distribution for  $M_T > 950 \text{ GeV}$ . The tighter  $M_T$  requirement provides additional suppression of the SM background, but the position of the endpoint is essentially unchanged.

Figure 3.9(a) would seem to indicate that the trileptons +  $b$ -jets search does not have useful sensitivity to the STC model. However, the events in this figure are restricted to the  $N(b) \geq$



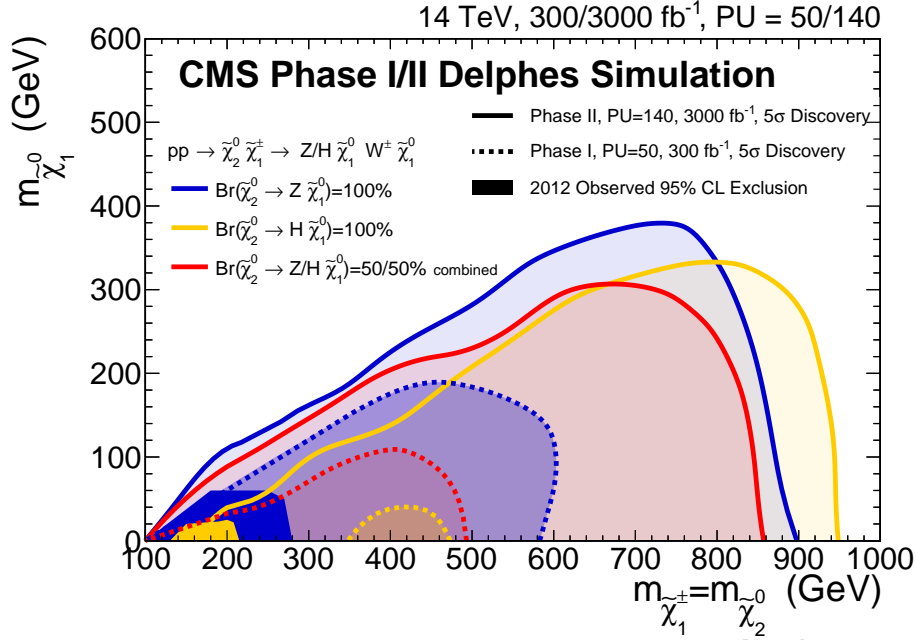


Figure 3.11: Search for  $\tilde{\chi}_2^0 \tilde{\chi}_1^\pm$  production in the  $W^\pm Z + E_T^{\text{miss}}$  and  $W^\pm H + E_T^{\text{miss}}$  final state. The excluded regions are shown in the simplified model parameter space of  $m(\tilde{\chi}_1^0)$  vs.  $m(\tilde{\chi}_1^\pm) = m(\tilde{\chi}_2^0)$  for various assumptions. In such plots, the mass of the produced particle (or particles) is generally shown on the x-axis, while the mass of the LSP is shown on the y-axis. As a consequence, the excluded region is bounded by the decreasing production cross section on the right, but by the decreasing  $E_T^{\text{miss}}$  as one approaches the diagonal.

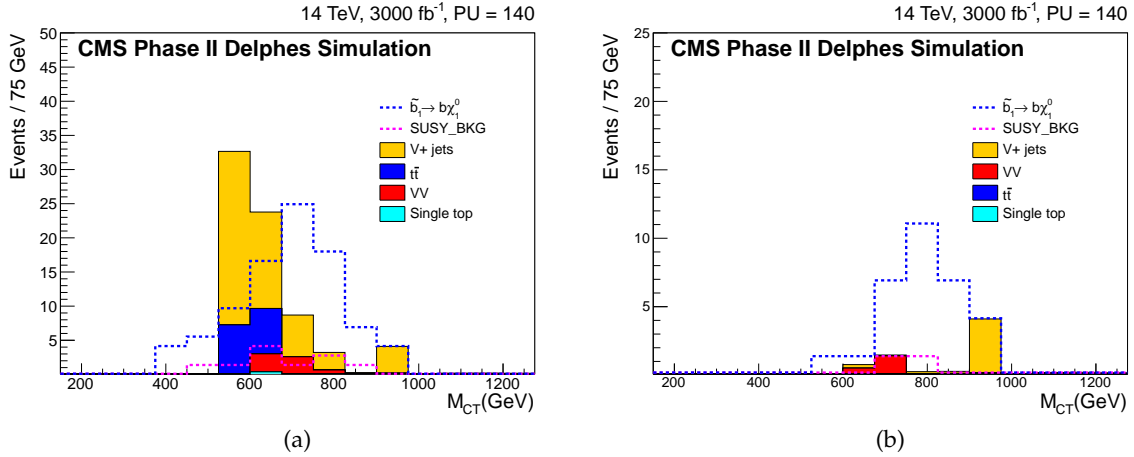


Figure 3.12: All-hadronic  $\tilde{b}_1$  search. Distributions of the  $M_{CT}$  variable for (a)  $M_T > 750$  GeV and (b)  $M_T > 950$  GeV. The endpoint of the  $M_{CT}$  distribution is a function of  $m(\tilde{b}_1)$  and  $m(\tilde{\chi}_1^0)$ .

4 multiplicity bin. For the  $N(b)=2-3$  bin, the search reaches  $5\sigma$  sensitivity at an integrated luminosity around  $500 \text{ fb}^{-1}$ . The absence of a signal in the four b-tagged jet bin indicates that the gluino is heavy, so that gluino pair production (with  $\tilde{g} \rightarrow \tilde{b}\tilde{b}$ ) is suppressed.

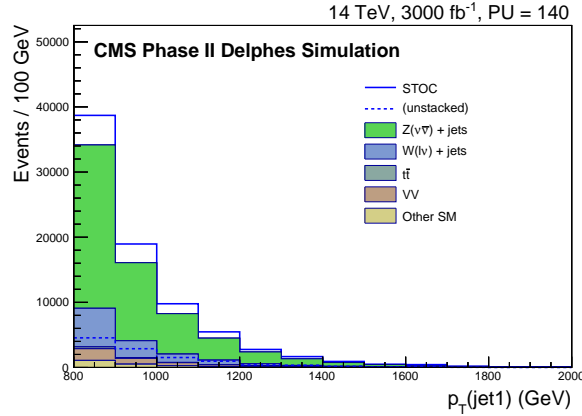


Figure 3.13: Monojet-like search in STOC: distribution of the  $p_T$  of the leading jet in the event.

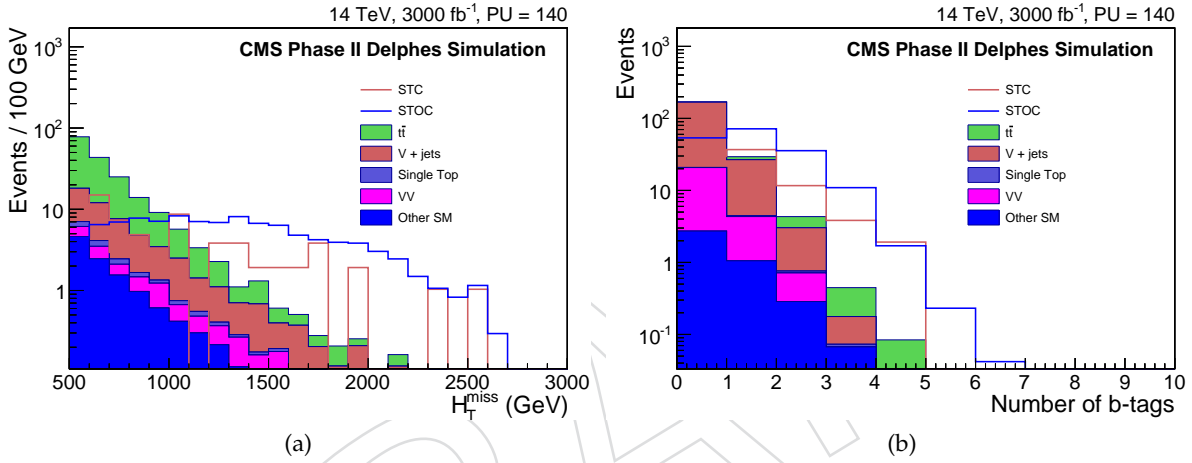


Figure 3.14: All-hadronic  $H_T - H_T^{\text{miss}}$  search. (a) Distribution of  $H_T^{\text{miss}}$  for the SM backgrounds and the STC and STOC signal contributions. (b) Distribution of the number of b-tagged jets after all other selection requirements.

### 3.3.2.3 Results for the stop coannihilation model (STOC)

As noted in Sec. 3.3.1.3, the very low top squark mass in STOC leads to a large cross section for direct top-squark production. Because the  $\tilde{t}$  and  $\tilde{\chi}_1^0$  are nearly mass degenerate, however, the  $\tilde{t}_1$  is effectively invisible, and the process is not accessible experimentally without the presence of an additional jet from initial-state radiation. Figure 3.13 shows the distribution of the leading jet  $p_T$  from the analysis of the monojet-like final state, where an excess above the SM background is evident. The analysis must determine the very large, irreducible background contribution from  $Z(\rightarrow \nu\bar{\nu}) + 1$  jet. The significance of this signal approaches  $5\sigma$  with  $3000 \text{ fb}^{-1}$ .

### 3.3.3 SUSY conclusions

If a SUSY spectrum of any kind emerges from the LHC program, we will have finally broken through to beyond-the-SM physics in accelerator experiments. The studies presented here show that a broad range of searches will provide a way to discover and characterize what may be a very complex spectrum.

Several conclusions arise from these studies.

- The breadth of the SUSY search program implies that all of the capabilities of the

CMS detector are required.

- For some of the measurements, the performance of the degraded detector will severely compromise the program.
- For natural SUSY models, which are designed to address the gauge hierarchy problem, discovery in one or more of the favorable search signatures could occur within the first  $300 \text{ fb}^{-1}$  of data taking. Sensitivity to other key signatures, however, can require  $33000 \text{ fb}^{-1}$ .
- Because a given search channel provides only limited information, interpreting any observed excesses in terms of a specific particle spectrum will be a major challenge.
- To map out the properties of a particle spectrum, it is essential to have the full pattern of results that will be obtained at the highest integrated luminosities.

The program of SUSY measurements will provide critical information needed to address the issues of dark matter, the unification of forces, and the gauge hierarchy program.

### 3.4 Non-SUSY searches for DM

Astrophysical evidence for Dark Matter (DM) is one of the few existing hints for physics beyond the SM. The LHC can study many different types of possible interactions between quarks and DM. The LHC operating in high luminosity mode will deeply probe search regions with very low DM masses ( $\lesssim 10 \text{ GeV}$ ) and regions of phase-space where direct-detection experiments are blind because of their irreducible background from neutrino-nucleus scattering.

To illustrate the physics sensitivity in HL-LHC, we consider the case where a W boson recoils against a pair of DM particles, which, following the W boson decay, can yield an event containing an electron plus missing transverse energy. Events containing a single electron with  $p_T > 100 \text{ GeV}$  (following the Run-I single-electron trigger threshold) are selected if the ratio of the electron  $p_T$  and missing transverse energy is in the range  $0.4 < p_T / E_T^{\text{miss}} < 1.5$  and the angle  $\Delta\phi$  between both objects is larger than 2.5. There is an implicit  $M_T$  cut due to the lepton  $p_T$  requirement. The discriminating variable is the transverse mass,  $M_T$ , calculated from the electron  $E_T$  and missing transverse energy which is caused by the neutrino.

As a special feature, this channel is sensitive to different couplings of DM to up- and down-type quarks - parametrized by  $\xi$  - and may provide an explanation for inconsistent results from DD experiments. The production can be modeled with an effective field approach (EFT) where no specific assumptions apply to the mediator other than it being very heavy (labeled  $\Lambda$ ), or with a simplified model where the mediator mass and its couplings (and therefore its width) are specified. For high mediator masses as will be probed at the HL-LHC, both models yield comparable description and sensitivity[50]. The strong dependence of the  $M_T$  spectrum and the signal efficiency on the value of  $\xi$  is due to the significantly different  $M_T$  shape.

Discovery would reach EFT scales (or mediator masses) of  $\Lambda > 3.7$  for  $\xi = -1$  and  $\Lambda > 2.6$  for  $\xi = +1$ . Fig. 3.15 illustrates the situation for the monolepton channel in terms of 95% C.L. exclusion limits using interpretations in the framework of a simplified model. The mediator is modeled as a  $Z'$ -like particle with either vector or axial-vector coupling, a fixed minimal width of  $\Gamma_{\text{med}} = M_{\text{med}} / 8\pi$  and the mass range shown in Fig. 3.15. The monojet channel, which is often used for comparison (see ref [51] for projection) relies on an initial state radiation jet for tagging the event, which, along with missing transverse energy due to the produced DM-pair, leads to a final state of jet +  $E_T^{\text{miss}}$ . This channel is expected to profit from the improved jet performance discussed in Chapter 9. Given its kinematics it cannot be sensitive to interference but

its sensitivity is comparable to the monojet channel with constructive interference ( $\xi=-1$ ). The projected performance of the monojet channel is based on Run-I performance projected to larger accumulated statistics [51]. The monojet sensitivity in Fig. 3.15 is again based on DELPHES simulation and illustrates the situation for the maximum sensitivity corresponding to constructive interference ( $\xi=-1$ ) and the narrow width indicated in the legend. The hadronic W boson channel should reach even further but has not been studied in this framework.

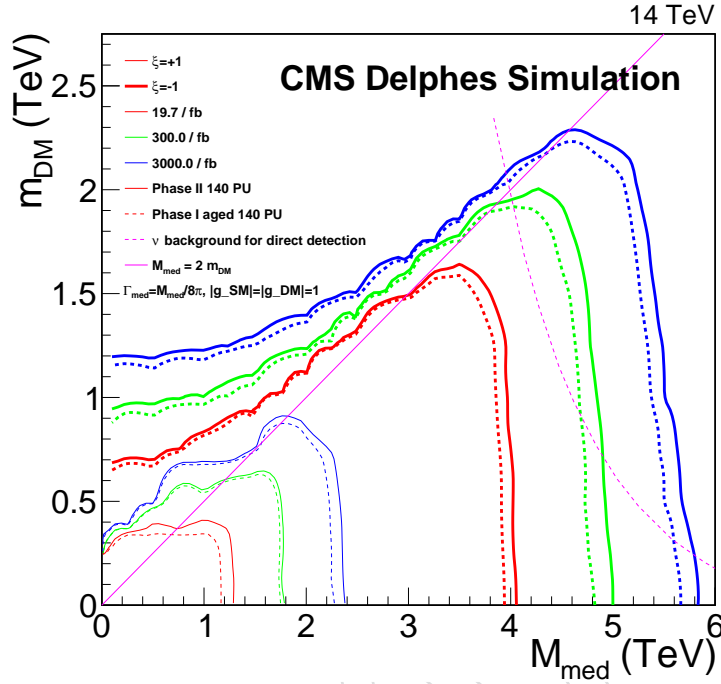


Figure 3.15: Dark matter reach of the monojet channel as a function of the DM mass and mediator mass for the two extreme cases of  $\xi = \pm 1$ .

### 3.5 Studying Rare SM Processes

In the SM, EWSB occurs via the Higgs mechanism where the vector bosons acquire mass through their coupling to the Higgs field. Along with that, the scattering amplitudes of longitudinally polarized W or Z vector bosons ( $V_L V_L \rightarrow V_L V_L$ ) are unitarized through the presence of the Higgs boson. The presence of the additional interactions with the Higgs introduces the new terms in the scattering cross section calculation that remove the divergence at high energy that would otherwise occur. If the SM is a low-energy effective theory as is widely thought, the unitarization of VBS could only be partially caused by the Higgs boson, and the full regularization would happen via some new physics that intervenes at a high energy scale. With the large dataset of the HL-LHC, VBS processes may be studied in order to determine whether this is the case.

This section presents studies of the potential of the CMS experiment at the HL-LHC in this context. The VBS signal will be accessible when two quarks from the beams emit vector bosons, which in turn interact with each other. The two quarks get deflected from the beam direction and originate jets of hadrons in the detector, called tag jets, which clearly identify this category of events. In this way the HL-LHC is used as a vector boson collider, providing access to purely electroweak processes in a high energy regime. Any new resonances in the VBS invariant mass

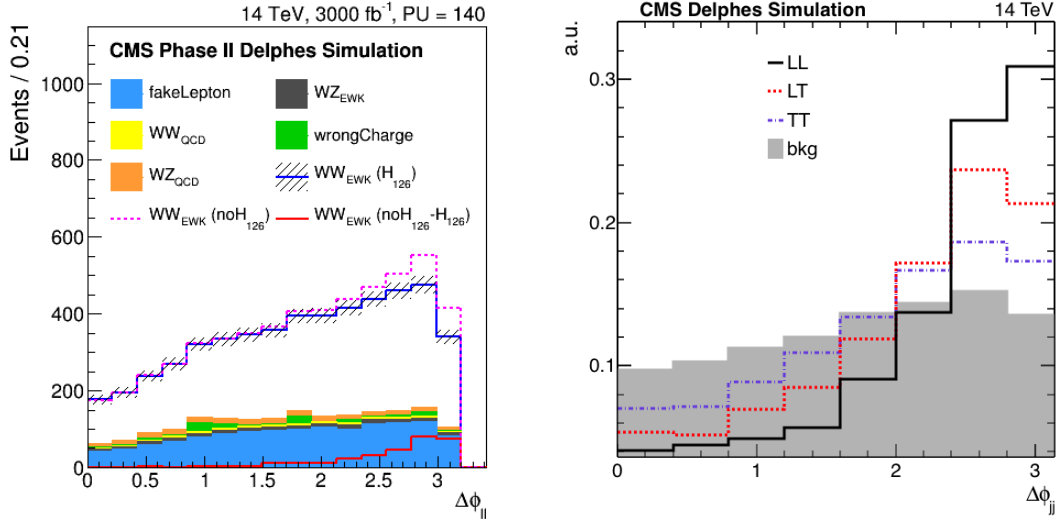


Figure 3.16: Left, the  $\Delta\phi$  between the two final state charged leptons for the same-sign WW scattering, after the VBS selections, for positive muons in the non-aged Phase-I scenario. Right, an example of the expected differences for polarized scatterings in the WZ analysis.

spectrum, or a deviation of data from the SM expectations would provide direct evidence of the scale at which new physics enters into play.

The small cross section of this electroweak process, and the large background due to the production of vector boson pairs and two hadronic jets via strong interactions (irreducible background), make this analysis very challenging. In the high-pileup environment of the HL-LHC, background contamination can also arise from cases where hadronic jets are by mistake identified as leptons (j-l misidentification rate), or jets from different events overlap and mimic the features of tagging jets. The new central tracking system will be beneficial in reducing the j-l misidentification rate and its extension to  $|\eta| = 4$ , together with a new radiation resistant and highly granular forward calorimeter will reduce the contamination of jets from pileup events and allows to trigger on these events with higher efficiency.

For this study, only fully leptonic decays of the vector bosons are selected, each decaying into electrons or muons, in order to avoid the ambiguities due to the separation of the tag jets from the vector boson decay products. Projections are presented for same-sign WW scattering and WZ scattering, which are among the final states that are most sensitive to new physics.

The signal and irreducible background have been generated at leading order with the PHANTOM [52] and MADGRAPH [53] generators, while the detector response has been simulated parametrically with DELPHES [16]. Detector effects specific to these analyses, such as the j-l misidentification rate have been evaluated with the detailed simulation of the CMS detector, accounting for the specific conditions of each upgrade design. Anomalous couplings are studied in the frame of the EFT formalism [54] implemented in the MADGRAPH generator, while non-unitarized scenarios are described by the PHANTOM program. Final-state leptons are considered if they are found in the detector with  $|\eta| < 2.4$  and  $p_T > 20$  GeV, and jets have to satisfy  $|\eta| < 4.7$  and  $p_T > 30$  GeV.

The performance of the analysis has been evaluated for each final state, after applying rectangular selections on  $m_{jj}$  and  $\Delta\eta_{jj}$  exploiting the typical VBS signature. Figure 3.16, left, shows the  $\Delta\phi$  between the two final state charged leptons for the WW scattering, after the VBS selections. The signal is represented by the red line, stacked over the backgrounds. The dashed purple line

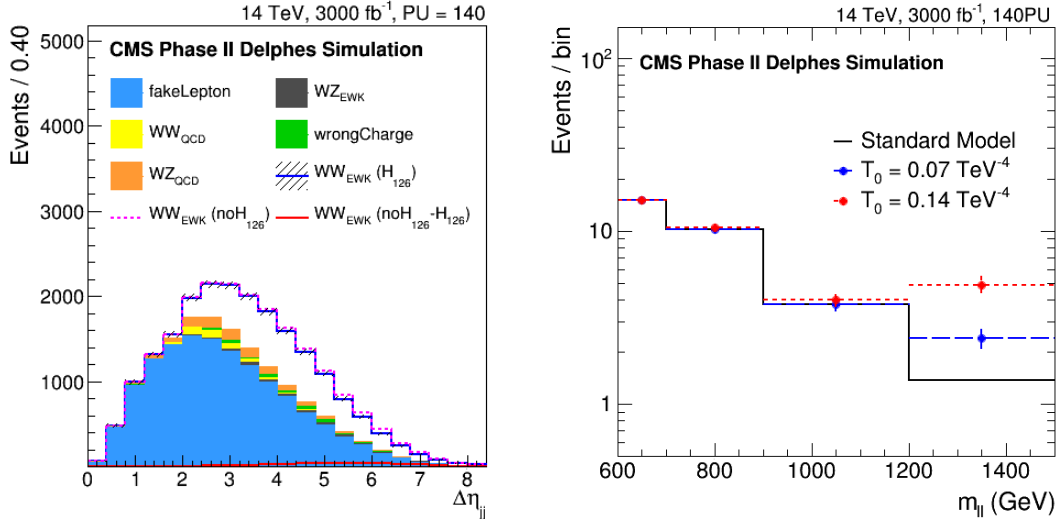


Figure 3.17: Left, the differences in shape of  $\Delta\eta_{jj}$  between the two final state jets for the WW scattering, before the VBS selections, for signal and background. Right, the  $m_{\ell\ell}$  distribution at the end of the analysis chain, still for the WW scattering, showing the expectation for the SM case and for a signal hypothesis with enhanced anomalous couplings.

shows the different expectation in the extreme case when the Higgs boson does not play any role in the WW scattering unitarization (no-Higgs scenario). The difference of the Higgsless scenario and the SM one is used as signal on top of the SM itself, as an indicator of the analysis sensitivity to models where the Higgs boson performs a partial unitarization of VBS. The distributions on the right-hand side show the expected differences for polarized scatterings in the WZ analysis. In this case the  $V_L V_L \rightarrow V_L V_L$  signal is searched for on top of the background and the other scattering components. Figure 3.17, on the left, shows the  $\Delta\eta_{jj}$  between the two final state jets for the WW scattering, before the VBS selections. On the right, the  $m_{\ell\ell}$  distribution is shown at the end of the analysis chain, still for WW scattering, showing the expectation for the SM case and for a signal hypothesis with enhanced anomalous couplings.

### 3.5.1 Same-sign WW scattering

Besides the two VBF tagged jets, the final state of this channel is composed of the two same-sign charged leptons from the W bosons decay (either electrons or muons) and missing transverse energy due to the undetected neutrinos. The irreducible background is well kept under control, as clearly visible in Figure 3.16, left. After kinematic selection requirements, the expected number of events after  $3000 \text{ fb}^{-1}$  of integrated luminosity are reported in Table 3.1 for the HL-LHC.

Scenario	WW			WZ	
	Signal	I-bkg	R-bkg	Signal	I-bkg
HL-LHC 140 PU	3155	374	1662	1100	1513

Table 3.1: The expected number of events, after the analysis selections, for same-sign WW and WZ final states. Signal, irreducible background (I-bkg) and reducible one (R-bkg) are reported. Each line corresponds to a possible scenario for the CMS operations.

After  $3000 \text{ fb}^{-1}$  of integrated luminosity the total cross section of the EWK production of same-sign W bosons plus two jets will be determined with a precision at the order of 5%, after accounting for the main expected sources of uncertainty.



Fig. 3.18 shows the expected 95% CL contour for the BSM  $S_0$  and  $S_1$  parameters. Since the deviations from the SM happen in high energy tails of the distributions, and the results approach the systematic limit already, the difference between the scenarios is typically at the order of 10%.

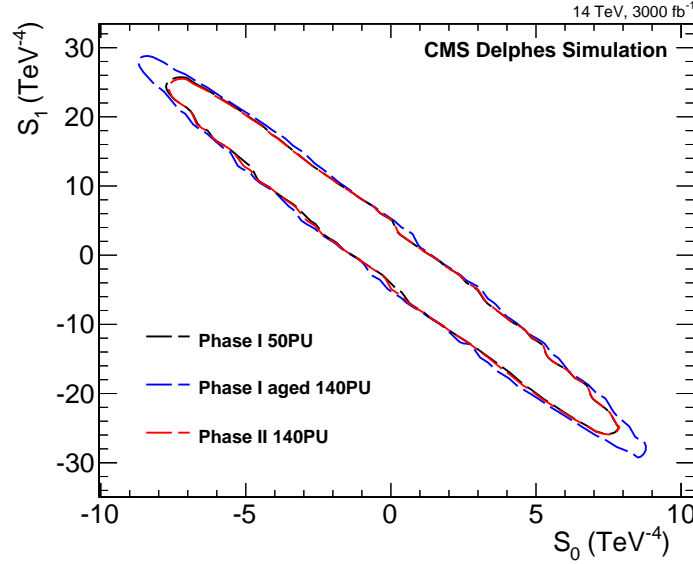


Figure 3.18: Expected 95% CL 2D contour for the  $S_0$  and  $S_1$  parameters (right).

### 3.5.2 WZ scattering

This channel searches for three charged leptons in the final state, where two of them have opposite sign, same flavor and an invariant mass compatible with the one of a Z boson. The undetected neutrino generates missing transverse energy, and the longitudinal component of its momentum can be determined by requiring it to be produced, together with the charged lepton, in the decay of a W boson. The background due to Drell–Yan events plus jets, when one jet is wrongly identified as a charged lepton. The high-granularity forward calorimeter technology and the tracker extension play a crucial role in identifying and removing pileup jets. Besides the VBS selections of  $m_{jj} > 600$  GeV and  $\Delta\eta_{jj} > 4.0$ , events should have a third reconstructed and identified lepton, with  $|\eta| < 2.4$  and  $p_T > 20$  GeV. Same-flavor, opposite-sign lepton pairs should not have a mass consistent with the Z boson mass within 6 GeV and have at least 20 GeV of mass. After these conservative requirements, the expected number of events after  $3000 \text{ fb}^{-1}$  of integrated luminosity are reported in Table 3.1 for the various possible upgrade scenarios considered.

These studies demonstrate that the upgraded CMS detector will provide the improvement in the physics performance necessary for the verification of the EWSB in a model-independent way with respect to the detailed description of the low-mass Higgs resonance.

### 3.5.3 B physics

The HL-LHC will open new possibilities for B physics measurements, especially the study of rare processes like the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays. On the other hand, the high instantaneous luminosity and increased pileup will create demanding conditions for the trigger and offline reconstruction of B production and decay.

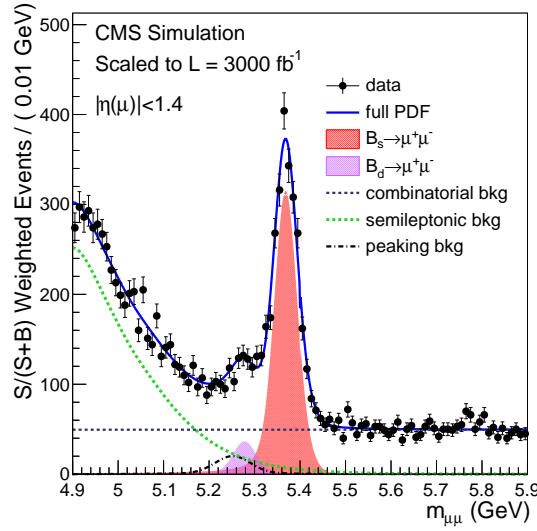


Figure 3.19: Projections of the mass fits to  $3000 \text{ fb}^{-1}$  of integrated luminosity, respectively assuming the expected performances of upgraded CMS detector. The plot is for the barrel only.

In this section, the B-physics potential of the upgraded CMS detector is illustrated by presenting a study of the  $B^0 \rightarrow \mu^+ \mu^-$  and  $B_s \rightarrow \mu^+ \mu^-$  decays. This work is based on the  $\mathcal{B}(B_s \rightarrow \mu^+ \mu^-)$  measurement published by the CMS collaboration with the LHC Run-I data [55], and focuses on the implementation of a prototype L1-trigger algorithm and on the estimate of the final analysis sensitivity.

The only way to build a L1 trigger for the  $B^0$  and  $B_s$  signal at the HL-LHC is by exploiting the track trigger architecture of the upgraded CMS detector. Without the L1 track trigger, inferior resolution and high pileup contamination would result in unsustainable event rates at the low thresholds needed for the measurement.

Figure 3.19 shows toy Monte Carlo projections of the  $B^0$  and  $B_s$  analysis results for the HL-LHC. The plot corresponds to barrel events and a total integrated luminosity of  $3000 \text{ fb}^{-1}$ . The  $\sigma \times \mathcal{B}$  predicted by the SM is assumed for  $B^0$  and  $B_s$ . The background models are taken from the  $\mathcal{B}(B_s \rightarrow \mu^+ \mu^-)$  measurement published by CMS with the LHC Run-I data. These results show that with HL-LHC the two peaks can be resolved due to the improved invariant mass resolution provided by the upgraded detector.

From these studies, the sensitivity for the observation of the  $B^0 \rightarrow \mu^+ \mu^-$  decay is expected to be  $\approx 6.8\sigma$ , while the branching fractions  $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$  and  $\mathcal{B}(B_s \rightarrow \mu^+ \mu^-)$  can be measured with a precision of 18% and 11% respectively for the HL-LHC. Their ratio  $\frac{\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)}{\mathcal{B}(B_s \rightarrow \mu^+ \mu^-)}$  can be measured with a 21% uncertainty for at HL-LHC.

### 3.6 Physics Requirements

This mapping between science goals and science requirements is graphically summarized in Table 3.2 below.

The flow down from the science requirements presented in the previous section to technical

Table 3.2: Science Goals  $\rightarrow$  Science Requirements Matrix

	Higgs Couplings	BSM Search/Study	DM	RareProcesses
b-tagging $\epsilon \sim 70\%$	✓	✓	–	✓
VBF tagging $\epsilon \sim Y\%$	✓	✓	✓	–
Momentum resolution $\sim X\%$	✓	✓	–	✓
Dijet resolution $\sim X\%$	✓	✓	–	–
Diphoton resolution $\sim X\%$	✓	✓	–	–
$E_T^{\text{miss}}$ resolution $\sim X\%$	✓	✓	✓	–
Trigger thresholds $\sim m_W/2$	✓	✓	✓	✓

requirements is summarized by the 3.3. We highlight those aspects of this program that are enabled by the proposed DOE scope of the US CMS HL-LHC upgrades to the CMS detector, namely the upgrade of the outer tracking system, the increase trigger bandwidth and the capabilities of the track trigger, the upgrades to the barrel and the replacement of the endcap calorimeter, and the extension of the pseudorapidity coverage in the muon system, the conceptual designs of which will be shown in subsequent chapters of this CDR.

### 3.6.1 Durable detector capable of high performance until $3 \text{ ab}^{-1}$ integrated

### 3.6.2 Efficient reconstruction for all Higgs Decays

### 3.6.3 Ability to disentangle different Higgs Production

### 3.6.4 Electroweak scale single-lepton trigger thresholds

### 3.6.5 Percent level lepton momentum resolution

### 3.6.6 Efficient b-tagging in central and forward Rapidities

### 3.6.7 Efficient VBF jet tagging and associated primary vertex resolution

### 3.6.8 Excellent missing transverse energy resolution

### 3.6.9 Pile-up mitigation such that LHC performance recovered at HL-LHC

- How the upgrades satisfy high level science requirements. Groundwork for this needs to be laid in each technical chapter.

Table 3.3: Science Requirements → Technical Requirements Matrix

	Requirement	b-tag	VBF tag	$p_T$	$m_{jj}$	$m_{\gamma\gamma}$	$E_T^{\text{miss}}$	Trigger threshold
FPIX								
	Tracking to $ \eta  = 4.0$	✓	✓	–	–	–	–	✓
	$\delta d_0 \sim X\%$	✓	–	–	–	–	–	✓
	$\delta\phi \sim X\%$	–	–	✓	–	–	–	✓
	$\delta z \sim X\%$	✓	✓	–	–	✓	✓	✓
Outer Tracker								
	$\delta p_T/p_T < 0.05$	✓	–	–	–	–	–	✓
	$\delta d_0 < 0.05$	✓	–	–	–	–	–	✓
	$\delta\phi < 0.0005$	–	–	✓	–	–	–	✓
	$\delta \cot(\theta) < 0.002$	✓	✓	–	–	✓	✓	✓
HCAL Barrel								
	< 50% light loss	–	–	–	✓	–	✓	✓
	Survive 3/ab	–	–	–	✓	–	✓	✓
ECAL Barrel								
	< 50% light loss	–	–	–	–	✓	–	✓
HGC								
	Eff. Pile-up < Run 2	–	✓	–	–	✓	✓	✓
Muons								
	L1 rate < Y kHz	–	–	–	–	–	–	✓
Trigger/DAQ								
	Tracking at L1	–	–	–	–	–	–	✓
	EB crystal-level data granularity	–	–	–	–	–	–	✓
	~ 95% muon efficiency at $p_T \sim 20$ GeV	–	–	–	–	–	–	✓
	L1 rate ~ 750 kHz	–	–	–	–	–	–	✓

DRAFT

## Chapter 4

# Outer Tracker - Editor Canepa 40 pages

This chapter contains first a brief introduction to the limitations of the Phase 1 tracking system (Sec. 4.1). After a general description of the upgraded tracker (Sec. 4.3 and 4.2), the focus will shift towards the R&D and production deliverables within the DOE scope (Sec. 4.5).

### 4.1 Limitations of the Phase 1 tracking system

The Run 1 tracking system was designed to operate efficiently up to an integrated luminosity of  $500 \text{ fb}^{-1}$ , and an average pileup of 20-30 collisions per bunch crossing. A sketch showing one quarter of the layout of the Run 1 tracker is presented in Fig. 4.1. The radial region below 200 mm is equipped with pixellated detectors. Beyond 200 mm, the Run 1 tracker features single-sided strip modules and double-sided modules composed of two back-to-back silicon strip detectors with a stereo angle of 100 mrad. Double-sided modules provide coarse measurements of the  $z$  and  $r$  coordinates in the barrel and endcap, respectively. The tracking system was designed to provide coverage up to a pseudorapidity of  $|\eta| \approx 2.4$ .

At the time of writing, the Run 1 pixel detector is being replaced by an upgraded detector referred to a “Phase 1” pixel detector [56]. The Phase 1 pixel detector is formed of 4 layers (instead of 3) in the barrel region and of 3 disks (instead of 2) in the end-cap region. The innermost layer is closer to the beam than the Run 1 pixel detector.

Studies of the expected performance of the Phase 1 tracking system as a function of the integrated luminosity have shown significant degradation beyond about  $1000 \text{ fb}^{-1}$ . The performance degradation has been studied extensively and is documented in the Technical Proposal for the CMS Phase 2 Upgrade [57] and in the Scope Document [58].

Accumulated radiation damage in the Phase 1 pixel sensors reduces the charge collection as well as the Lorentz angle. This first leads to lower charge sharing among neighbouring pixels and hence deteriorated spatial resolution, and eventually to reduced hit efficiency.

For the Run 1 strip tracker, the most prominent change of detector properties with irradiation is the increase of the sensor depletion voltage and that of leakage current. The latter can be only partially mitigated by lowering the operating temperature of the cooling system. It was shown that the double-sided strip modules could not be operated anymore for the nominal cooling temperature (Fig. 4.2).

A representative study is shown in Fig. 4.3 where the tracking efficiency is presented in the various scenarios. In addition to their intrinsic limitations, the Phase 1 tracking system restricts the CMS data acquisition to a maximum L1 accept rate of about 100 kHz, with an available latency of  $4 \mu\text{s}$  for the trigger decision. Operations at high luminosity will however be based on a substantial upgrade of the trigger system, with significantly higher rate capability and



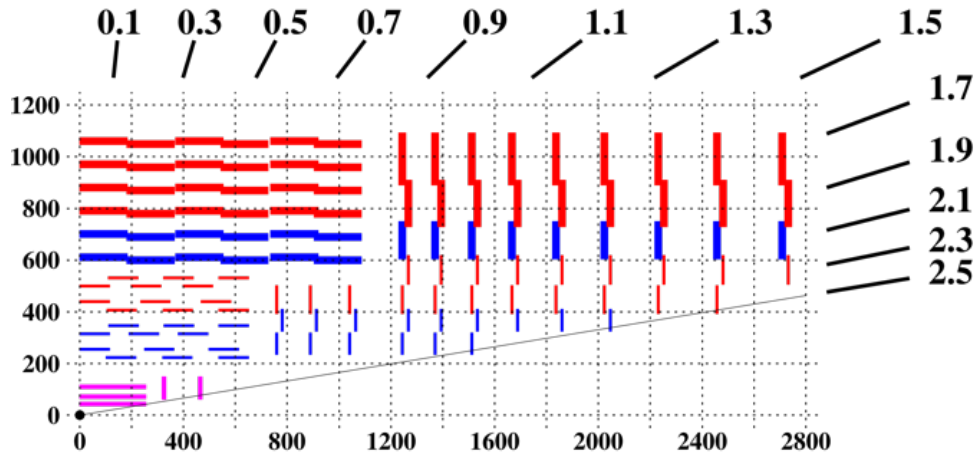


Figure 4.1: Sketch of one quarter of the Run 1 CMS tracking system. The radial region below 200 mm is equipped with pixellated detectors. Beyond 200 mm, the outer tracker features single-sided strip modules (red segments) and double-sided modules composed by two back-to-back silicon strip detectors with a stereo angle of 100 mrad (blue segments). Double-sided modules provide measurements of the  $z$  and  $r$  coordinates in the barrel and end-cap, respectively. The four inner tracker barrel layers (six outer tracker barrel layers) are referred to as TIB (TOB). The three inner forward disks (nine end cap disks) on each side are referred to as TID  $D_i$  (TEC  $D_i$ ). **FIXME:** This plot may be replaced, waiting for final version to be included in the TDR.

longer latency.

To maintain the excellent physics performances achieved during Run 1 and Run 2 under the HL-LHC conditions, the entire Phase 1 tracking system (Run 1 tracker and Phase 1 pixel detector) will be replaced with a detector providing enhanced radiation tolerance ( $\times 10$ ), increased granularity ( $\times 5$ ), and extended  $\eta$  coverage. The upgraded tracker, *i.e.* Phase 2 tracker, will consist of an Inner Tracker (IT) based on silicon pixel modules, and an Outer Tracker (OT) assembled from silicon modules with strip and macro-pixel sensors. The upgraded outer tracker will provide input to the L1 trigger at 40 MHz. This upgrade will ensure that CMS adapt to the challenges of the LHC accelerator and fully exploits its physics potential.

## 4.2 Requirements for the Phase 2 tracker system

The main requirements for the Phase 2 tracker can be classified into design, technical, and physics requirements. The first class includes the requirements that the tracker must satisfy to operate efficiently at the HL-LHC and within the upgraded CMS. The technical requirements are those ensuring that the tracker provides precise and robust measurements for pile-up values as high as 140-200. Finally the physics requirements are those necessary to achieve high performance in object reconstruction and identification. The latter are described in Sec. 3.6 while the design and technical requirements are summarized below:

- Design Requirements:

- Radiation tolerance: The upgraded tracker must be able to operate efficiently up to a target integrated luminosity of  $3000 \text{ fb}^{-1}$ . This requirement must be fulfilled without any maintenance intervention for the Outer Tracker, while for the Inner Tracker it is envisaged to replace modules

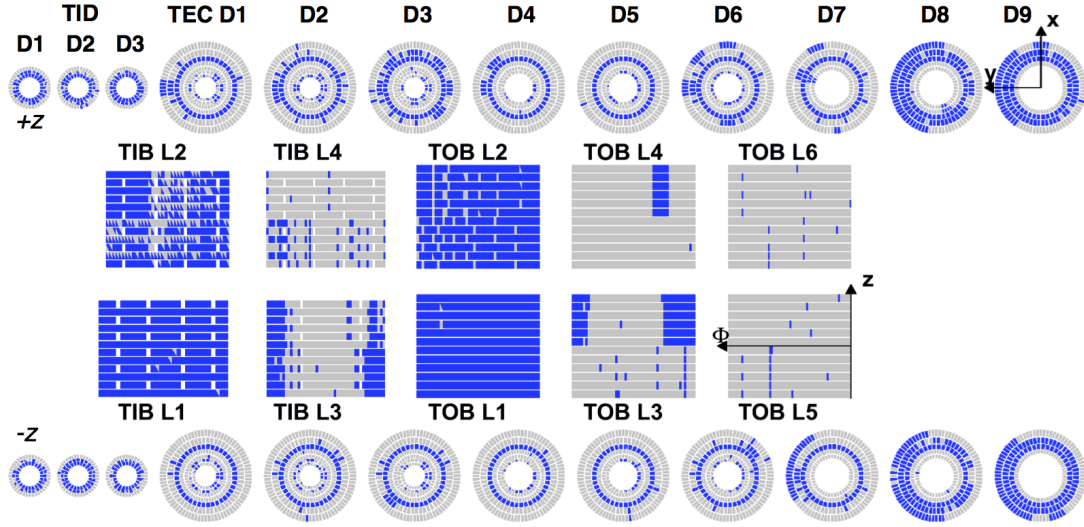


Figure 4.2: Map of non-functional modules (in blue) after an integrated luminosity of  $1000 \text{ fb}^{-1}$ , for the achievable minimum coolant temperature of  $-20^\circ\text{C}$ . Almost all the stereo modules in the barrel (Inner Barrel layers 1 and 2 and Outer Barrel layers 1 and 2), as well as in the endcap (rings 1, 2 and 5), are no longer operational. The evolution of the leakage current of the tracker sensors is predicted by a detailed model that takes into account the estimated luminosity profile, the position and size of each module, the expected particle fluence at specific module locations (obtained from FLUKA simulations and the expected temperature versus time scenario that includes annealing periods). The model also implements a map of the efficiency of the module thermal contacts derived from data.

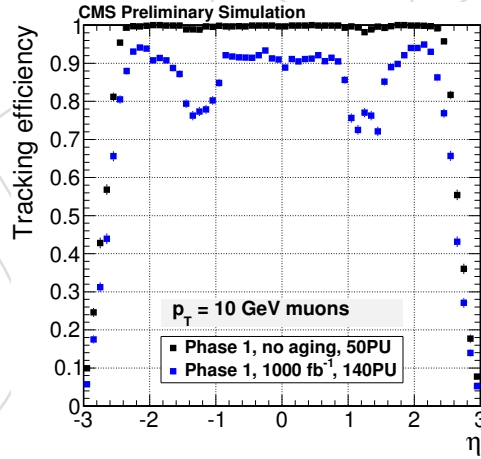


Figure 4.3: Reconstruction efficiency for  $p_T = 10 \text{ GeV}$  muons as a function of pseudorapidity for the Phase-I tracker before and after the tracker has been aged by an equivalent integrated luminosity of  $1000 \text{ fb}^{-1}$ .

and other elements in the innermost regions, as they accumulate substantial radiation damage. Detailed FLUKA [59, 60] simulations have been performed to estimate the radiation exposure of the different detector regions. The exposure is about one order of magnitude higher than the specification used to design the Run 1 tracker [61, 62]. It reaches a 1 MeV neutron equivalent fluence per  $\text{cm}^2$  of  $2 \times 10^{16} \text{ n}_{\text{eq}}/\text{cm}^2$  in the innermost pixel regions. The particle fluence strongly depends on  $r$ , while the varia-

1895  
1896  
1897  
1898  
1899  
1900  
1901

tion with  $z$  is very moderate (Fig. 4.4).

- Compliance with the L1 trigger upgrade: The Outer Tracker is expected to provide input to the upgraded track trigger within a latency of  $12.5 \mu\text{s}$ . Furthermore, the new electronic system (with the exception of the analogue ASIC front-ends) has to be fully digital to cope with the substantially larger channel count. (Sec. 4.4.3).

- Technical Requirements:

- High granularity: occupancies as low as 1% are required to achieve efficient and robust pattern recognition and tracking. Target figures of 140 and 200 collisions per bunch crossing are used to benchmark the performance of the detector.
- Optimal layout of layers: optimal layout along with high granularity is essential to ensure efficient and robust pattern recognition, efficient track reconstruction, and suppression of track mis-reconstruction probability. It is also needed to provide excellent momentum and impact parameter resolutions.
- Reduced material in the tracking volume: minimal amount of material is required to suppress multiple scattering, improve the performance of the calorimeters in the global event reconstruction, and reduce the photon conversion probability.

- Physics requirements include:

- excellent momentum and impact parameter resolution;
- efficient pattern recognition and tracking;
- efficient reconstruction of the vertex associated to the hard scatter and of the vertices associated to the pile-up interactions;
- efficient two-track separation;
- efficient reconstruction of secondary vertices;
- efficient reconstruction of converted photons;
- extended tracking acceptance.

**FIXME:** I assume that Chris will place these physics constraints into the bigger picture, no need to repeat the motivation here.

### 4.3 Brief overview of the Phase 2 tracker system

The Phase 2 tracker, a quarter of which is shown in Fig. 4.5, contains the Inner Tracker and the Outer Tracker.

The Inner Tracker will be equipped with pixel modules. Pixel modules are formed of a 100-150  $\mu\text{m}$  thick pixel sensors ( $25 \times 100 \mu\text{m}^2$ , or  $50 \times 50 \mu\text{m}^2$ , or 3D ) bump-bonded to a readout chip based on 65 nm CMOS technology. The baseline Inner Tracker layout is shown in Fig. 4.5. The detector comprises a barrel part with four layers (referred to as Tracker Barrel Pixel Detector or TBPX), eight small double-discs per side (referred to as Tracker Forward Pixel Detector or TFPX) and four large double-discs per side (referred to as Tracker Endcap Pixel Detector or TEPX). In the TBPX the pixel modules are arranged in ladders, while in TFPX and TEPX the pixel modules are arranged in concentric rings. Each double-disc is physically made of two discs, which allows to mount modules onto four planes. Each disc is split into two halves, and these D-shaped structures are referred to as dees. The TEPX will provide the required luminos-

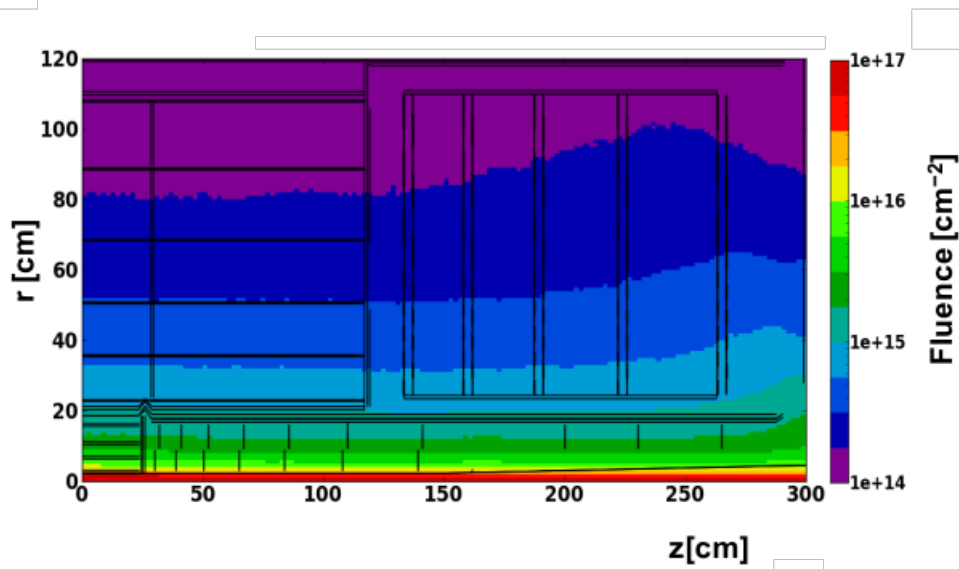


Figure 4.4: Integrated particle fluence in 1 MeV neutron equivalent per  $\text{cm}^2$ , for the Phase 2 tracker. The estimates shown here correspond to a total integrated luminosity of  $3000 \text{ fb}^{-1}$  of  $pp$  collisions at  $\sqrt{s} = 14 \text{ TeV}$ . **FIXME:** Plot needs to be updated. TID plot to be added. Update geometry overlay.

ity measurement capability by an appropriate implementation of the readout architecture. In total the pixel detector will have an active surface of approximately  $4.9 \text{ m}^2$ .

The Outer Tracker is populated by so-called  $p_T$  modules. The  $p_T$  modules are composed of two single-sided closely-spaced sensors read out by a common set of front-end ASICs, that correlate the signals in the two sensors and select hit pairs (referred to as stubs). The communication between the bottom and top sensors is made through folded flex circuits hosting the front-end electronics (Fig. 4.6).

Two classes of  $p_T$  modules are planned for the Outer Tracker: modules with two strip sensors (2-strip or 2S modules) and modules with a strip and a macro-pixel sensor (pixel-strip or PS modules). The strips in the 2S module have a length of about 5 cm, while those in the PS modules are 2.5 cm long. In the PS modules one of the two sensors is segmented into macro-pixels of 1.5 mm length, providing also a measurement of the  $z$  coordinate. The Outer Tracker is composed of six hermetic cylindrical “barrel” layers in the central region ( $|z| < 1200 \text{ mm}$ ) complemented on each side by five hermetic “endcap” double-discs located in the forward par t( $1200 < |z| < 2700 \text{ mm}$ ). Three sub-detectors are distinguished, as illustrated in FIG: the Tracker Barrel with PS modules, TBPS; the Tracker Barrel with 2S modules, TB2S; and the Tracker Endcap Double-Discs. The PS modules are deployed in the first three layers of the Outer Tracker, in the radial region of 200–600 mm, while 2S modules are deployed in the outermost three layers, in the radial region above 600 mm.

**FIXME:** will point to the figure once it contains TPBS ... Being prepared for the TDR

It is important to note that the PS modules in the inner three layers are installed at large  $z$  with increasing tilt angle. The optimal tilt angle is that keeping the sensors perpendicular to the trajectories of particles originating from  $z = 0$ . The non-tilted (“flat”) and tilted regions are

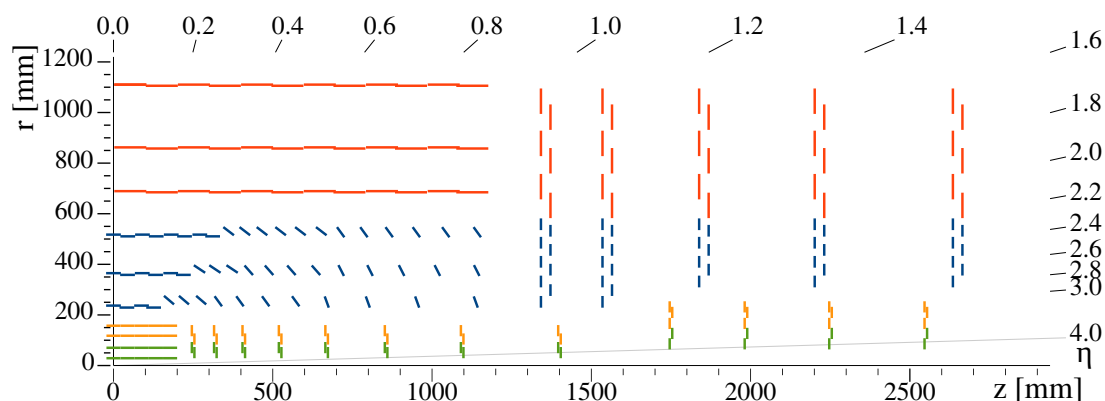


Figure 4.5: Sketch of one quarter of the Phase 2 tracker layout in  $r$ - $z$  view. In the Inner Tracker the green lines correspond to pixel modules made of two readout chips and yellow lines to pixel modules with four readout chips. In the Outer tracker the blue lines correspond to PS modules, while red lines correspond to 2S modules. Details are provided in the text. **FIXME:** This figure will have to be replaced once more, as the number of rings in layer 1 of the TBPS increased to 12.

referred to as Flat TBPS and Tilted TBPS, respectively.

**The Inner Tracker is outside the scope of this document. In what follows the details about the Outer Tracker are provided.**

## 4.4 Overview of the Outer Tracker

### 4.4.1 Outer tracker sensors

Silicon strip and macro-pixel sensors are the primary detection devices of the Outer Tracker. They have to provide sufficient spatial resolution to separate close-by-track in high density environment. . At the same time they have to be sufficiently radiation tolerant to withstand the doses expected at the HL-LHC.

#### 4.4.1.1 Radiation hard material

A comprehensive campaign has been carried out to identify suitable silicon materials for the Outer Tracker at HL-LHC [63? , 64]. Different test structures and sensors were implemented on selected wafer materials and produced by a single vendor, Hamamatsu Photonics K.K. (HPK). The devices were then electrically characterised, irradiated with different fluences of neutrons and protons, and then characterised again.

In addition to the standard wafer material, namely float zone (FZ) and magnetic Czochralski (MCz) silicon, an other subspecies of FZ silicon called deep diffused float zone (ddFZ) was investigated and found suitable as well. It offers the possibility to restrict the active thickness of the sensor compared to its physical thickness (here  $320\ \mu\text{m}$ ). The sensors show the electrical advantages of thin sensors at costs which are similar to thick sensors. Concerning the contribution to the material budget, physically thinned sensors are preferred.

Close attention was paid to charge collection (CC), noise behaviour and strip isolation before and after irradiation. Other crucial parameters, which define the requirements for services like power and cooling, are the bias voltage needed to extract sufficient signal and the resulting leakage current.



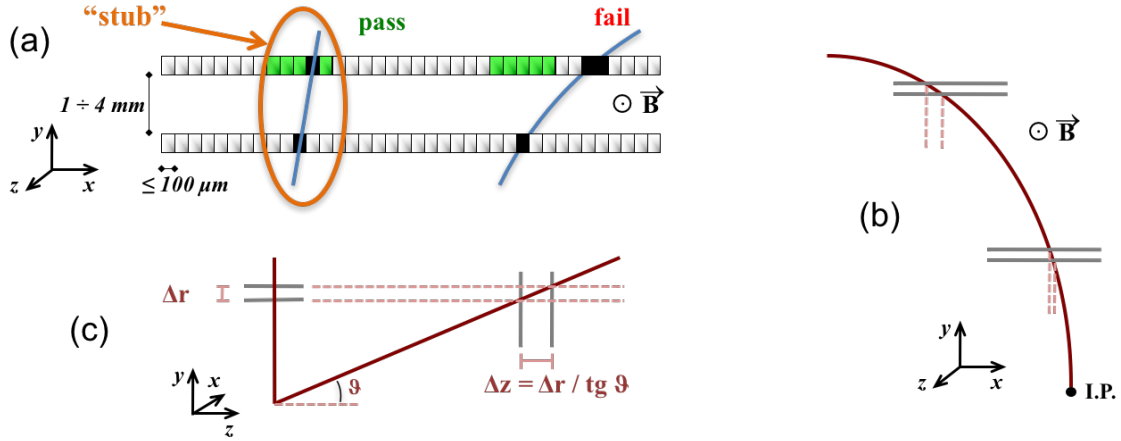


Figure 4.6: Illustration of the  $p_T$  module concept. (a) Correlation of signals in closely-spaced sensors enables rejection of low- $p_T$  particles; the channels shown in green represent the selection window to define an accepted stub. (b) The same transverse momentum corresponds to a larger distance between the two signals at large radii for a given sensor spacing. (c) For the endcap discs, a larger spacing between the sensors is needed to achieve the same discriminating power as in the barrel at the same radius. The acceptance window can therefore be tuned along with a few different values of sensor spacing to achieve the desired  $p_T$  filtering in different regions of the detector.

It was observed that sensors thinner than  $300 \mu\text{m}$  exhibit a lower full depletion voltage and a lower leakage current. As reported previously [65], p-in-n (p-type implants in an n-type bulk, collection of holes) sensors suffer from a higher degradation of the signal with irradiation than n-in-p (n-type implants in a p-type bulk, collection of electrons) sensors. Nevertheless, the signal extracted from  $200 \mu\text{m}$ -thick sensors after maximum fluence irradiation was found to be sufficient for both types. More influential for the choice of the sensor polarity was the observation that highly irradiated p-in-n sensors suffer from a significant non-Gaussian noise contribution, while this effect was not observed in n-in-p sensors. This noise could result in an irreducible rate of fake hits as large as the expected signal occupancy in the detector, disqualifying sensors with hole readout.

Sensors of n-in-p type require dedicated implantations to achieve proper isolation between neighbouring strips. Both investigated strip isolation techniques, p-spray and p-stop, ensure sufficient signal separation before and after irradiation.

#### 4.4.1.2 Design and specifications

Each 2S module will require two identical 2S strip sensors. These are AC-coupled strip sensors where each strip is segmented into two 5 cm long strips at a pitch of  $90 \mu\text{m}$ . At each end of the  $10 \times 10 \text{ cm}^2$  sensor 1016 strips are read out by eight chips, resulting in a total of 2032 channels. A bias ring surrounding the active area distributes the ground potential to each strip via polysilicon resistors. High voltage stability up to  $-800 \text{ V}$  is ensured using a single floating guard ring<sup>1</sup>. A highly doped  $p^+$  implant at the periphery ensures a well defined depletion volume and thus protects the sensor from damage due to wafer dicing. Alignment marks, strip

<sup>1</sup>The bias ring or grid of the sensor is connected to the same ground potential as the readout chips, while the high potential is applied to the backplane. For n-in-p type sensors the backplane has to be on negative potential to deplete the bulk. For the operation of the sensors itself only the potential difference is relevant. For the depletion voltage thus only absolute values are quoted in the remainder of this section.



Table 4.1: Most important specifications of the three sensor types used in the Outer Tracker.  
**FIXME:** these numbers have to be double checked.

Sensor name	Width [mm]	Length [mm]	Strip/Pixel		Quantity needed	Quantity with spares
			pitch	length		
2S	94.183	102.7	90 $\mu\text{m}$	50 mm	16448	18920
PS-s	98.14	49.16	100 $\mu\text{m}$	25 mm	5332	6130
PS-p	98.74	49.16	100 $\mu\text{m}$	1.446 mm	5332	6400

numbering and labels are implemented on this p+ implant. Each sensor will receive a unique identification tag encoded in scratch pads or laser fuses.

The PS module contains one strip sensor (PS-s sensor) and one macro-pixel sensor (PS-p sensor). The design of the PS-s sensor is similar to the 2S sensor except for outer dimensions and the strip geometry. With a size of  $10 \times 5 \text{ cm}^2$  and strips of 2.5 cm length it represents only half of the surface of a 2S sensor. The width of the PS-s sensor is 300  $\mu\text{m}$  smaller than that of the PS-p sensor at each end. During module assembly the alignment marks on the lower PS-p sensor are still visible from the top and can be used for precise alignment between the two sensors. On each front-end hybrid eight SSA chips match a row of 960 strips at a pitch of 100  $\mu\text{m}$ . The PS-p sensor is a DC-coupled pixel sensor with 1446  $\mu\text{m}$  long macro-pixels at a pitch of 100  $\mu\text{m}$ . Each macro-pixel is connected to a common bias grid using punch-through structures. The layouts of the bias ring, the guard ring and the periphery are again very similar to those of the strip sensors.

The size of the sensors in 2S modules is defined to optimally fit in one 6" silicon wafer, while for PS modules two sensors are obtained from one wafer. The reduced size (half-length) of PS modules is necessary to be able to cover the surface of the pixelated sensor with two rows of front-end (FE) ASICs bump-bonded to the sensor itself. At the same time, the shorter strips in PS modules are appropriate for the higher track density in the inner regions of the Outer Tracker, keeping the occupancy at acceptable levels. The choices of pitch of 90  $\mu\text{m}$  and 100  $\mu\text{m}$  are driven by limitations in the line density on the readout hybrid for the 2S module, and on the bump density of the C4 technology for the PS module. The strip sensors are wire bonded directly to the front-end hybrids, thus avoiding heavy pitch adapters.

The resistivity of the base material shall be larger than 3 k $\Omega$  cm to ensure that the initial full depletion voltage ( $V_{fd}$ ) remains below 300 V.

The most significant specifications of the sensor layouts are summarised in Tab. 4.1.

#### 4.4.1.3 Operational aspects

The minimum operation voltage for the sensors shall be at least 10% above the depletion voltage  $V_{fd}$  of each sensor. Regardless of the final choice of sensor thickness and resistivity, the initial  $V_{fd}$  will be smaller than 300 V, therefore an operation voltage at the sensor backplane of  $-400 \text{ V}$  will be sufficient during the first years of operation. A seed signal of 12,000 electrons can be extracted from 200  $\mu\text{m}$  thick sensors at a depletion voltage of 600 V even after exposure to a fluence of  $6 \times 10^{14} \text{ n}_{eq}/\text{cm}^2$ . Nevertheless, sensors, modules and HV power supplies are specified to support a maximum operation voltage of  $-800 \text{ V}$ , which allows for an additional boost in signal for modules which are most affected by a reduction of the signal-to-noise ratio.

The reverse bias current that a sensor draws after exposure to irradiation is described using the

current related damage factor<sup>2</sup> referred to as  $\alpha$ .

The value of  $\alpha$  is conservatively assumed to be  $6.16 \times 10^{-17} \text{ A cm}^{-1}$  at the end of the sensor lifetime. Using the maximum fluences of  $3.3 \times 10^{14} \text{ n}_{\text{eq}}/\text{cm}^2$  for 2S and  $1.0 \times 10^{15} \text{ n}_{\text{eq}}/\text{cm}^2$  for PS sensors, the maximum currents can be estimated. For PS (2S) sensors of an active thickness of  $200 \mu\text{m}$  a maximum current of about 1.2 mA (0.8 mA) is expected for a sensor temperature of  $-20^\circ\text{C}$ . This results in a maximum power dissipation of approximately 1 W per sensor, which has to be removed by cooling.

#### 4.4.1.4 Production and quality control

The production of the sensors will be contracted to qualified industrial partners. To mitigate the risks of a producer failing to fulfill its obligation, the production will be split between at least two companies, where each is able to produce the full amount of sensors. The qualification of suitable vendors is ongoing at the time of writing within the framework of a CERN Market Survey. The final selection of material (MCz, FZ or ddFZ), physical thickness, strip isolation technique (p-stop or p-spray) and production technology (6" or 8") will depend on the capabilities and costs offered by the selected companies.

Only sensors which can provide the expected performance over the full lifetime of the detector will be integrated into modules. To ensure that sensors are of sufficient quality, each has to comply to a detailed set of electrical specifications. The compliance with these specifications will be tested using the following quality control procedures:

- Vendor Quality Control (VQC): After production, each sensor shall be electrically characterised by the producer, and the measurement results shall be recorded in a database (CMS construction database). Only sensors in compliance with the electrical specifications will be shipped to CERN.
- Sensor Quality Control (SQC): One or two sensors per production batch<sup>3</sup> (corresponding to 5-10% of the full quantity) shall be fully characterised at a SQC center to qualify the full batch.
- Process Quality Control (PQC): Test structures produced on the same wafer as the sensors enable a faster and more detailed characterisation of the production process. Measurements on test structures from two to four wafers per batch (corresponding to 10-20% of the full quantity) shall be performed at a PQC center to qualify the full batch.
- Irradiation Tests (IT): A small sample of sensors and test structures (in the order of 1% of sensors and 5% of test structures) will be irradiated and then electrically characterised to ensure the expected radiation hardness for all sensors from the series production.

Only good sensors from batches which have passed VQC and PQC will be used for module assembly. This production scheme follows closely the scheme employed in the sensor production for the Run 1 tracker.

<sup>2</sup>Bulk damage due to hadron irradiation will increase the leakage current significantly. The rise of the bulk current is proportional to the fluence, as described as:  $\Delta I = \alpha \Phi_{\text{eq}} V$  where  $\Phi_{\text{eq}}$  is the fluence in 1 MeV neutron equivalents per  $\text{cm}^2$ ,  $V$  is the depleted volume and  $\Delta I$  is the resulting increase in leakage current.

<sup>3</sup>A size of about 20 wafers per batch is assumed. If the batch size used by a vendor differs, the sample size has to be adjusted accordingly.

#### 4.4.2 Outer tracker modules

One of the essential requirements of the Outer Tracker is the capability to provide tracking information to the L1 trigger decision, implying that the tracker will send out self-selected information at every bunch crossing. Furthermore CMS plans to enhance the first level trigger rate from 100 to 750 kHz and to increase the latency from 3.2 to 12.5  $\mu$ s. These data taking conditions translate into requirements for the tracker front-end electronics. Such functionality will rely upon local data reduction in the front-end electronics, in order to limit the volume of data that have to be sent out at 40 MHz. This is achieved with the  $p_T$  modules mentioned in Sec.4.3. These modules are capable of rejecting signals from particles with transverse momentum lower than a given  $p_T$  threshold. A threshold of around 2 GeV corresponds to a data volume reduction of roughly one order of magnitude, which is sufficient to enable transmission of the stubs at 40 MHz, while the other signals are stored in the front-end pipelines and read out when a trigger signal is received. For a pitch of about 100  $\mu$ m between silicon strips (or macro-pixels), sufficient  $p_T$  resolution can be achieved for radii as small as 200 mm in a barrel geometry, thanks to the 3.8 T magnetic field of CMS. The concept is therefore applicable in the Outer Tracker, and limited in angular acceptance to about  $|\eta| < 2.4$ .

The  $p_T$  module concept requires the readout electronics that performs stub finding to be connected to both the top and the bottom silicon sensor of a module. In order to implement the connectivity between the upper and lower sensors with reliable and affordable technologies, the two halves of each module are read out independently by the front-end hybrids on the two ends (referred to as FEH). The power and the electrical-to-optical conversion is hosted on service hybrids (SEH).

##### 4.4.2.1 Implementation of $p_T$ modules

The main parameters of 2S and PS modules are summarized in Tab. 4.2 while Fig. 4.7 (Fig. 4.8) shows assembled modules (an exploded view of modules). Sketches of the connectivity between sensors and readout hybrid circuits are shown in Fig. 4.9.

Table 4.2: Main parameters of the 2S module and the PS module.

2S module		PS module	
separation 1.8, 4.0 mm		separation 1.6, 2.6, 4.0 mm	
2 $\times$ 1016 strips:	5 cm $\times$ 90 $\mu$ m	2 $\times$ 960 strips:	2.5 cm $\times$ 100 $\mu$ m
2 $\times$ 1016 strips:	5 cm $\times$ 90 $\mu$ m	32 $\times$ 960 macro-pixels:	1.5 mm $\times$ 100 $\mu$ m

In both modules the desired gap between the two sensors is achieved by gluing them onto spacers made of aluminium (Al) / carbon fibre (CF) composite (Al-CF). The spacers provide mechanical support and efficient heat removal. They exhibit high thermal conductivity in particular in-plane, and a coefficient of thermal expansion well matched to silicon.

Different values of the gap between the two sensors of a module are needed, in addition to a programmable acceptance window in the front-end ASICs, in order to implement a coherent  $p_T$  filtering in the whole Outer Tracker volume. For this reason 2S modules are realized in two versions, with 1.8 mm and 4.0 mm gaps between the centres of the active volumes of the sensors, and PS modules are realized in three versions, with 1.6 mm, 2.6 mm and 4.0 mm gaps. Table 4.3 summarizes the number of modules per type and version in the different sub-detectors.

The front-end hybrids are realized in a flexible (flex) technology, and they are laminated onto CF stiffeners. The hybrids are folded around spacers matching the thickness of the assembly of the two sensors. The folding and lamination steps will be performed in industry. In all hybrids,

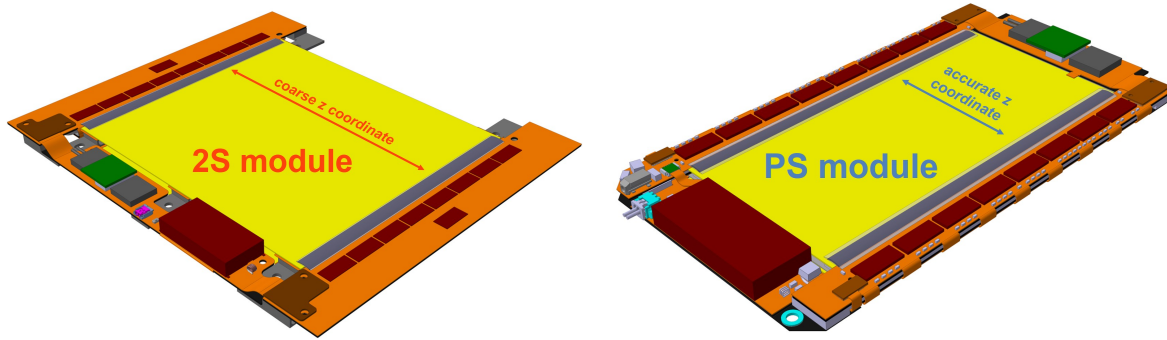


Figure 4.7: The 2S module (left) and PS module (right) of the Outer Tracker. Shown are views of the assembled modules. The 2S module includes the silicon sensors (yellow), the front-end hybrids hosting 8 chip each (orange), and one service hybrids hosting HV, LV, and the electrical-to-optical converter. The PS module includes the silicon sensors (yellow), the front-end hybrids hosting 8 chip each (orange), one service hybrids hosting the LV and HV, and one hybrid hosting the electrical-to-optical converter. The bottom of the PS module (not visible here) contains the pixellated detector bump-bonded to the readout electronics. Details are given in the text, and in the following sections.

Table 4.3: Summary of module types and versions in the Outer Tracker. **FIXME:** Does not yet include the change to twelve rings in TBPS layer 1.

Module type and version		TBPS	TB2S	TEDD	Total per variant	Total per type
2S	1.8 mm	0	4464	2824	7288	8224
	4.0 mm	0	0	936	936	
PS	1.6 mm	826	0	0	826	5332
	2.6 mm	1462	0	0	1462	
	4.0 mm	548	0	2496	3044	
Total		2836	4464	6256	13556	

the ASICs are bump-bonded onto the flex circuit.

One 2S front-end hybrid carries eight CMS Binary Chips or (CBC) reading out the strips of the top and bottom sensors at one sensor end, plus the Concentrator Integrated Circuit (CIC), which serves as interface between all the CBCs of the hybrid and the readout link. Both the CBC and the CIC are described in more details in Sec. 4.4.3. The role of the CIC is mainly to aggregate and serialize the data of the readout chips and to distribute clock, trigger and control signals to them. One PS front-end hybrid houses eight Short Strip ASICs (SSAs, described in Sec. 4.4.3) reading out the strip sensor, and the same CIC as used for 2S hybrids. In the PS modules the pixellated sensor is bump-bonded to its readout chips and form the MaPSA. All the front-end chips implement binary readout. In order to fully exploit the achievable hit position resolution, one extra bit is added to the hit address, such that in the case of clusters with an even number of fired channels the coordinate is set in the centre of the cluster (“half-strip resolution”). The chip logic also implements an offset correction, separately programmable for eight groups of strips in each chip, in order to form an unbiased estimate of the particle transverse momentum in the front-end electronics. Such programmable parameter can absorb also possible small offsets in the relative positions of the two sensors originating from limited

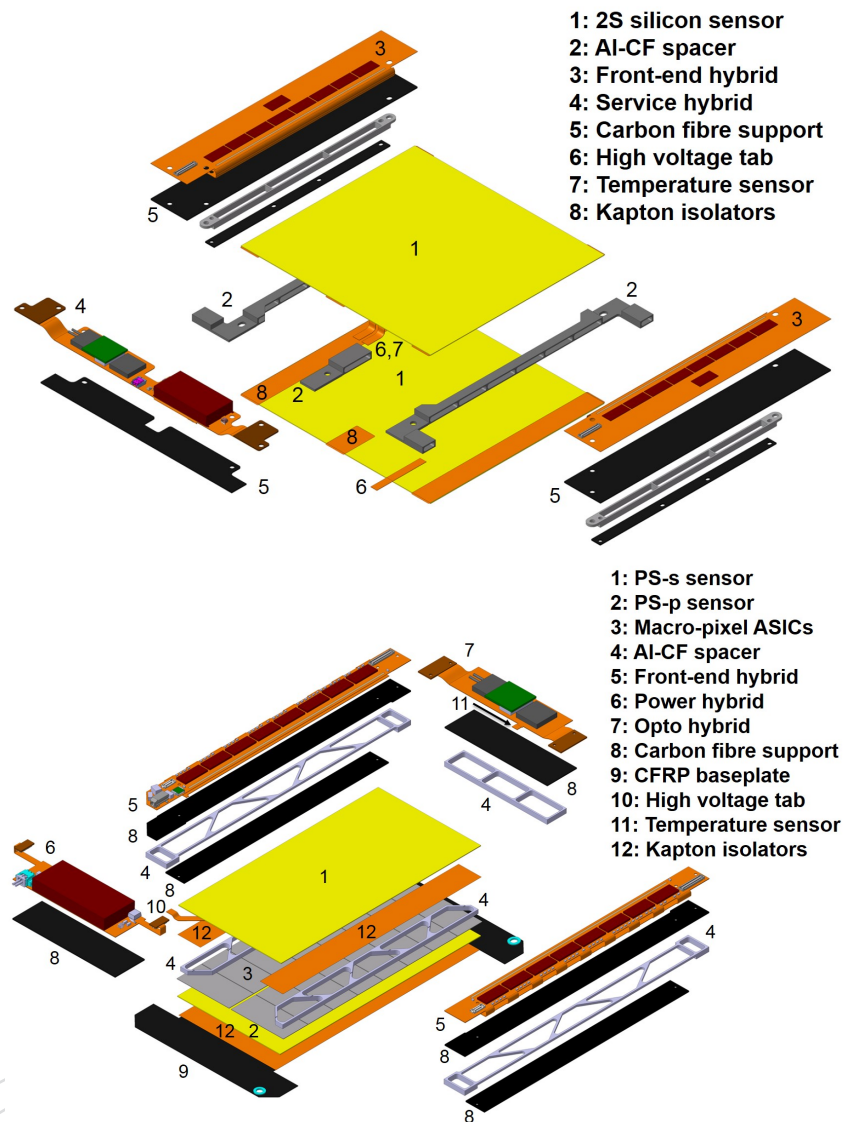


Figure 4.8: The 2S module (left) and PS module (right) of the Outer Tracker. Shown are details of the module parts.

precision in the mechanical assembly. However, a relative tilt between the two sensors cannot be corrected for in the chip logic, hence an optimal  $p_T$  estimate in the Outer Tracker modules translates into a requirement for the maximum angular tilt between the sensors.

The auxiliary electronics for powering and optical readout is integrated on service hybrids realized in the same flex technology as the front-end hybrids. The service hybrids are also laminated onto stiffeners. In 2S modules one single service hybrid is located on one end of the sensor assembly. In PS modules, due to the reduced width of these modules, powering and readout functionalities are implemented in two separate circuits, located on the two ends of the sensor assembly.

The 2S modules are mounted on the supporting structures with small screws at the sides of the sensor spacers. These mounting points provide cooling to the entire module, including its readout electronics. In PS modules the large front-end ASICs are thermally coupled to the



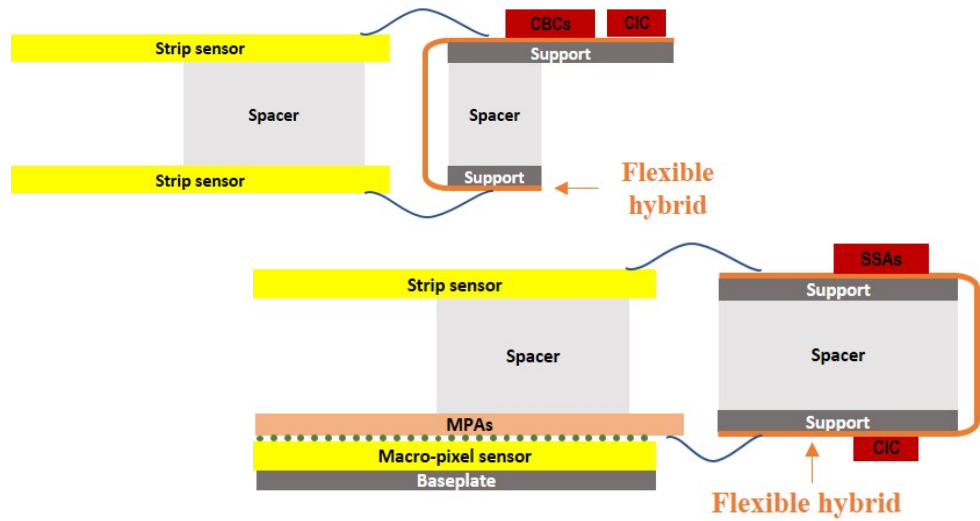


Figure 4.9: The 2S module (top) and PS module (bottom) of the Outer Tracker. Shown are sketches of the front-end hybrid folded assembly and connectivity. The description of the various electronics components can be found in Sec. 4.4.3.

pixelated sensor through the bumps, hence a large-surface cooling contact is required. The FE ASICs for the pixelated sensors, called MPAs, are described in more details in Sec. 4.4.3. For this reason the module is built on a baseplate made of carbon fibre reinforced polymer (CFRP). This baseplate is glued onto a cold surface on the supporting structure, in which the cooling pipe itself is embedded. The sensor spacers provide the thermal path from the strip sensor to the pixelated sensor, and the whole module is cooled through the baseplate.

#### 4.4.2.2 Module thermal performance

The thermal performance of both module designs and different sensor separations was studied with finite element analyses (FEA). The power consumptions of the front-end ASICs used in the calculations were estimated based on prototypes and simulations, whereas the sensor power consumption is calculated from the expected worst case fluence for a given module type and sub-detector, as obtained from FLUKA [59, 60] simulations.

The power consumption of the module is dominated by that of the electronics components even at the end of lifetime of the silicon sensors. The power consumption assumed for the front-end components in the FEA is  $\sim 7.8$  W and  $5.4$  W for the PS and 2S module, respectively.

The thermal performance of a module and its corresponding cooling structure is characterised in terms of the temperature at which the module undergoes thermal runaway when the coolant temperature is increased. For this purpose a temperature dependency is introduced to the sensor power consumption by scaling the value calculated at room temperature according to [66], and applying the resulting temperature dependent heat load to the individual sensor elements in the FEA model. The temperatures at which thermal runaway occurs are summarised in Tab. 4.4 for the nominal heat loads after receiving  $3000 \text{ fb}^{-1}$  of integrated luminosity. The cooling system is designed such that the coolant at the first module in a cooling loop, corresponding to the warmest module along the loop, has a temperature of  $-33^\circ\text{C}$ . For all module types thermal runaway is expected to occur at temperatures at least 11 K above the coolant temperature, indicating that sufficient margin is available.



Table 4.4: Summary of expected thermal runaway turn-on temperatures for the various module types and sub-detectors. Calculations were performed using the expected nominal heat loads after  $3000 \text{ fb}^{-1}$  of integrated luminosity. **FIXME:** Numbers for PS modules to be added when available.

Module type	Sub-detector	Thermal runaway turn-on temperature [ $^{\circ}\text{C}$ ]
2S 1.8 mm	TB2S	-21.5
2S 1.8 mm	TEDD	-19.9
2S 4.0 mm	TEDD	-21.7
PS 1.6 mm	TBPS	-X.YZ
PS 2.6 mm	TBPS	-X.YZ
PS 4.0 mm	TBPS	-X.YZ
PS 4.0 mm	TEDD	-X.YZ

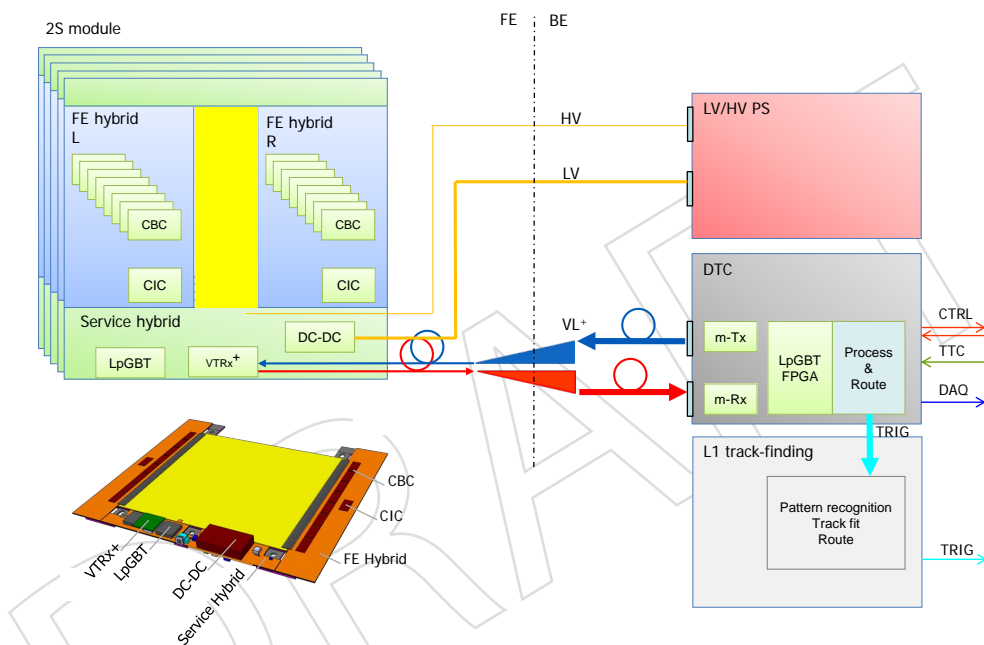


Figure 4.10: Electronic system block diagram, exemplified for the 2S module, together with a labelled sketch of the module. Details are provided in the text. On the data-trigger-control board (DTC), m-Tx and m-Rx are multi-channel transmit and receive optical modules. The L1 track-finding block is covered in **FIXME:** refer to the trigger chapter

### 4.4.3 Outer tracker electronics

The electronic system is designed to deliver trigger data at 40 MHz and L1 readout data with high efficiency at L1 accept rates up to 750 kHz, and to cope with latencies up to  $12.5 \mu\text{s}$ . Its block diagram is shown in Fig. 4.10.

#### 4.4.3.1 Architecture

As briefly introduced earlier, two electronic hybrids physically surround the sensor module (2S or PS). In 2S modules the two FE hybrids are wire bonded to the right (R) and left (L) strips of the two silicon sensors. In the case of the PS module, two hybrids are wire bonded

to the left and right strips of the strip sensor plus to a macro-pixel sub-assembly (MaPSA) that integrates the pixelated sensor with its readout chips. Data generated by eight FE chips (CMS Binary Chip (CBC) in the 2S module, Short Strip ASIC (SSA) and Macro-Pixel ASIC (MPA) in the PS module) is buffered, aggregated and formatted by the concentrator ASIC (Concentrator Integrated Circuit (CIC)) that acts as a data hub to the service hybrid.

The service hybrid hosts all services to/from the counting room: bidirectional optical data transfer (VL+), low voltage power (LV) and high voltage bias (HV). It connects to the two FE hybrids either via two miniature connectors or through wire bonds made at the same time as those to the sensors. Due to the factor two difference in width between the 2S and PS modules, the service hybrid must be split in two in the case of the PS module. There will be one hybrid on each PS module end, one for powering and the other for data transfer. The low power GBT (LpGBT [67]) ASIC serialises/deserialises data sent to/received from the VTRx+ (Versatile TRansceiver plus) optoelectronic transceiver. It also acts as I<sup>2</sup>C master of the module (controlling, monitoring and configuring the FE ASICs) and contains additional monitoring functionality that will be used to check the environment and operational parameters of the module.

Two different technologies have been selected for the FE ASICs: 130 nm and 65 nm CMOS. The choices were driven by the need to optimize the required performance (speed, power, density) while limiting the development and production risks and costs.

At the back-end (BE), the Data, Trigger and Control board (DTC) sends and receives data to/from 72 modules. This board is a custom developed ATCA (Advanced Telecom Computer Architecture) blade based on commercial FPGAs and multi-channel optoelectronic transceivers. It processes three data streams to/from the detector: data acquisition (DAQ), trigger (TRIG) and timing & control (TTC & CTRL, where TTC stands for Timing, Trigger and Control). The DAQ stream refers to the full event data sent out upon reception of a L1-accept trigger signal, the TRIG stream refers to the data sent to the L1 trigger at bunch crossing rate, and the TTC & CTRL stream includes the clock and trigger signals as well as configuration commands.

#### 4.4.3.2 On-module electronics

The Outer Tracker modules are finely segmented: 2032 strips on each side of 2S modules and 1920 strips plus 30720 macro-pixels for PS modules. **FIXME:** check numbers!!! This results in large amounts of data being processed by the front-end chips, which must be appropriately funnelled with minimal efficiency loss to a single optical link. The data generated on-module are aggregated in two steps, as mentioned earlier.

Each CBC processes data from 254 strips (127 bottom and 127 top sensor strips), identifies clusters of programmable maximum width, performs top to bottom correlations over configurable windows and generates high  $p_T$  stub data at bunch crossing (BX) rate. In addition, it provides unsparisified binary readout data at L1 rate. The CBC analogue front-end stage is expected to contribute less than 1000 electrons of noise to signals from strips with a capacitance of 8 pF. A peaking time of approximately 20 ns and a return to baseline within 50 ns allow the chip to process hits in consecutive bunch crossings without dead time.

The CBCs connect to the FEH via 725 bumps on a 250  $\mu\text{m}$  pitch. They are flip-chipped and soldered to the FEH during the same reflow step as all other passive components on the circuit. The CBCs exchange data with their neighbours to identify clusters spanning across chip boundaries, and send out trigger data and L1-accept readout data to the CIC. The data flow, shown schematically in Fig. 4.12, is organised in two separate paths: L1 readout (DAQ) and

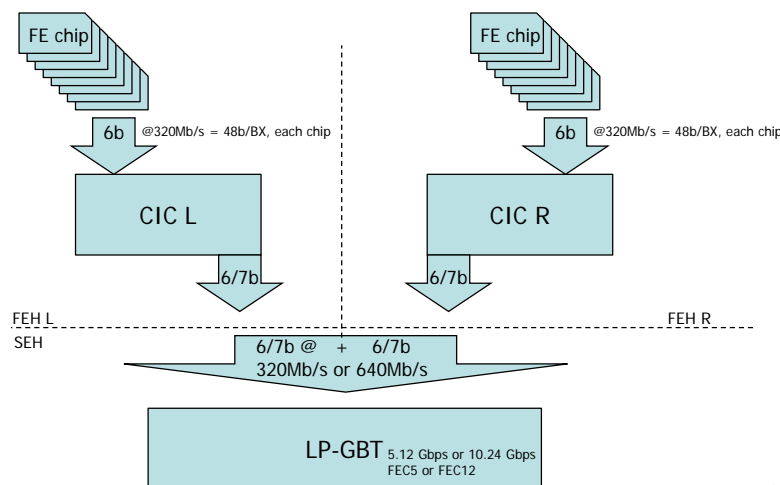


Figure 4.11: Illustration of the data flow at the modules' front-end. In the 2S system the LpGBT is used in 5 G/FEC12 mode (i.e. 6 b at 320 Mb/s from each CIC). In the PS system the LpGBT is used in 5 G/FEC5 or 10 G/FEC5 mode (i.e. 7 b at 320 Mb/s or 640 Mb/s from each CIC).

2238 Trigger (TRIG). At a data transfer rate of 320 Mb/s, each CBC chip sends 1 bit of DAQ and 5 bit  
 2239 of TRIG data to the concentrator every 3.125 ns. This bandwidth is compatible with transferring  
 2240 up to three trigger stubs from each CBC every BX, and sending unparsified readout data  
 2241 from each CBC pipeline up to an average 750 kHz L1-accept rate.

2242 The CIC chip buffers, sparsifies and aggregates the data from each CBC and funnels it to the  
 2243 service hybrid that provides a limited 1.92 Gb/s bandwidth capacity to each module side, i.e.  
 2244 6 bits at 320 Mb/s per FE Hybrid. This limited bandwidth is shared optimally by profiting from  
 2245 the statistical distribution of hits in time and space.

2246 The readout data is sparsified in the concentrator ASIC. One bit (out of 6 bits) is used by the  
 2247 DAQ channel out of the concentrator. Sparsification implies however that the DAQ channels of  
 2248 different modules will run asynchronously. Synchronous unparsified operation will however  
 2249 remain possible by selecting the corresponding concentrator operating mode, but only up to a  
 2250 100 kHz L1-accept rate.

2251 Trigger data is sent out of the concentrator (5 bits out of 6 bits at 320 Mb/s) in block synchronous  
 2252 mode, where stub data is aggregated over a fixed number of bunch crossings, hence smoothing  
 2253 the module occupancy fluctuations in time. Each block is 8 BX long and synchronous to all  
 2254 other TRIG blocks in the system. Stubs produced by any CBC on the FEH during this 8 BX  
 2255 window are accumulated in the TRIG block, up to a maximum of 16 stubs. Eventual additional  
 2256 stubs are discarded. This block transfer scheme allows to optimally share the limited LpGBT  
 2257 bandwidth in time (i.e. over eight consecutive 25 ns BXs, i.e. 320 bits), while upper-bounding  
 2258 the latency of the TRIG channel to 8 BXs.

2259 Simulations indicate that this approach is > 99% efficient even under high pileup conditions

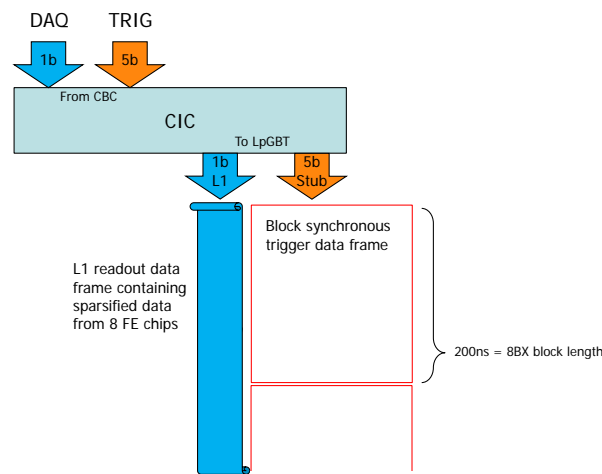


Figure 4.12: Illustration of the CIC data flow and formatting, exemplified for the 2S module.

(studied in samples with 140 pileup events plus four top quarks). As already mentioned above, some level of configurability will be included in the concentrator to make the system adaptable to different operation conditions.

Due to the pixelated nature of the PS module's bottom sensor, two front-end chips (both in 65 nm CMOS technology) are being developed for the PS system: one for the strip sensor (SSA) and one for the macro-pixel sensor (MPA). The correlation of strip and macro-pixel hits to build stubs takes place in the MPA chip.

The SSA ASIC sits on the FEH, much like the CBC in the 2S module case. It processes the strip sensor signals, and sends sparsified cluster data to the corresponding MPA chip at BX rate. The SSA analogue front-end stage features a simulated 24 ns peaking time and a 200 electron rms equivalent noise contribution for a strip capacitance of 4-5 pF.

The MPA ASIC is bump bonded to the macro-pixel sensor. Due to the large pixel size (1.446 mm x 100  $\mu$ m), a standard bump pitch of 200  $\mu$ m can be used, relaxing the assembly requirements. A total of 16 MPA chips are bumped to the macro-pixel sensor, resulting in an assembly (MaPSA, macro-pixel sub-assembly) with 30 208 bump bonded macro-pixels per module (1888 sensor bumps per MPA). The MPA analogue front-end stage features a peaking time of 24 ns and an equivalent noise contribution of 200 electron rms (260-280 fF pixel capacitance), very similar to the SSA expected performance. The MPA processes and sparsifies the hits from each macro-pixel. It correlates the bottom macro-pixel sensor hits with the data received from the SSA strips and builds stubs. Connections to and from the FEH are through wire bonds. As in the 2S module case, the CIC located on the FEH buffers and aggregates the stub and cluster data received from the MPAs, and sends them to the SEH. The transfer scheme and data formats are very similar to those used in the 2S module, allowing the same concentrator chip (with different configuration) to be used for both module types.

At a data transfer rate of 320 Mb/s, each MPA chip sends 1 bit of DAQ and 5 bits of TRIG data to the concentrator every 3.125 ns. This bandwidth is compatible with transferring up to five trigger stubs from each MPA every 2 BX, and sending all sparsified readout data from each MPA pipeline up to a 750 kHz L1-accept rate with negligible loss.

The limited bandwidth available from the serializer would unacceptably clip the amount of data which the CIC can pass on to the DTC if the same LpGBT configuration as for 2S modules would be used. Hence, the level of LpGBT forward error correction is decreased from FEC12 to FEC5 in the PS module case (up to five consecutive bits in error can be corrected instead of 12) to increase the number of available bits in the frame. From the concentrator, 1 bit of DAQ and 6 bits of TRIG data are sent to the LpGBT. For barrel layers 2 and 3 and disk rings 4 to 9 a transfer rate of 320 Mb/s is used, while 640 Mb/s is used for layer 1 and rings 1 to 3. The CIC is thus a dual rate chip which is configured to match the LpGBT operating data rate (either a low-speed/low-power 5.12 Gb/s raw data rate or a high-speed 10.24 Gb/s raw data rate). In low-speed mode, up to 17 stubs can be sent out from each CIC per 8 BX, while up to 35 stubs can be passed per 8 BX in the high speed mode.

Preliminary simulation results (derived from samples with 140 pileup events plus four top quarks) indicate that DAQ data can be transferred efficiently to the backend up to L1 rates of 750 kHz with a gradual degradation of performance above this rate (occasional buffer overflows, especially in the first barrel layer). TRIG stubs from particles with  $p_T$  above 2 GeV can be transferred with efficiencies  $> 99\%$  for the second and third barrel layer. For the first layer, due to the combination of higher particle density and reduced stub  $p_T$  resolution within the module, 99% efficiency can be achieved for particles with  $p_T$  above about 3 GeV, with a moderate loss of efficiency at the lower end of the  $p_T$  acceptance.

Timing is distributed on-module from an LpGBT 320 MHz E-port clock. In parallel, the data-channel of this E-port transmits to all front-end chips the trigger and soft-reset signals (Fast Reset, Trigger, Test Pulse Trigger, Orbit Reset) as well as a unique, repetitive BX-synchronous reference pattern which allows to unambiguously phase align the module to the 40 MHz BX clock.

The LpGBT ASIC includes slow control and monitoring features which can be accessed from the backend via a reserved set of bits in the data frame. Three I<sup>2</sup>C masters (each operating at 1 Mb/s) allow to control and configure all front-end chips on the two FEHs (left and right) and the VTRx+ on the SEH. In addition, dedicated reset lines allow to hard-reset groups of chips if necessary (CICs, CBCs, SSAs, MPAs). For monitoring purposes, five analog and three digital signals are measured, giving a good overview of the module operating conditions: sensor temperature, supply voltage and currents, input optical power, plus two (2S) or three (PS) status bits from the two or three DC-DC converters powering the module.

#### 4.4.3.3 Data formats, flows and rates

In what follows the two data streams flows introduced earlier are described in more details.

The DAQ data is a stream of frames responding to the L1-accept trigger signal. One frame corresponding to one L1-event is generated by each CIC chip upon reception of a L1-accept signal. Each frame aggregates the hit cluster data received from all FE chips associated to this CIC and belonging to that particular event. The frame size is flexible and depends on the number of hits in the module. DAQ streams from different modules are thus asynchronous across the detector and must be reorganized by the DTC. At an average L1-rate of 750 kHz, the sustainable capacity per module (two CICs) and per L1-accept event is 54 hits for 2S modules, 48 hits for PS modules in TBPS layers 2-3 and TEDD rings 4-9, and 102 hits for PS modules in TBPS layer 1 and TEDD rings 1-3. Peak capacity can of course be much higher than these values since the system accepts consecutive triggers, but cannot be sustained.

The TRIG data is a sequential stream of eight bunch crossings (8 BX) long blocks. Each block



contains the TRIG data aggregated from eight FE chips associated to the generating CIC. All TRIG data blocks generated by all modules are synchronous across the detector. The system capacity per module (two CICs) and per 8 BX block amounts to 32 stubs for 2S modules, 34 stubs for PS modules in TBPS layers 2-3 and TEDD rings 4-9, and 70 stubs for PS modules in TBPS layer 1 and TEDD rings 1-3. This is a maximum capacity value as TRIG data, contrary to DAQ data, is generated at BX rate.

#### 4.4.3.4 On-module services

Each CMS tracker module is connected to its individual set of services and is thus an autonomous element in the system. The service hybrid hosts all connections to the BE and associated electronic components: the DC-DC converters generating the necessary voltages for the FE electronics (1.0 V and 1.25 V) and optoelectronics (2.55 V); miniature HV connectors for the sensor bias lines; the LpGBT chip serializing/deserializing data to/from the BE, controlling the FE ASICs, distributing clock and trigger signals and providing local monitoring; the VTRx+ optoelectronic transceiver converting all up/downstream data to optical/electrical.

DC-DC converters, LpGBT and VTRx+ components are all derived from common developments for the upgrades of the HL-LHC experiments.

#### 4.4.3.5 Off-detector electronics

When compared to the FE, the level of effort devoted to off-detector electronics has only started to rise recently. This is due to the fact that most developments will be based on commercial off-the-shelf parts with limited environmental qualification needs. Moreover, significant technological evolution can be expected in the coming years and a timely development schedule will allow CMS to fully profit from it. Therefore only a conceptual design is sketched here, sufficient to present a probable implementation scenario.

The existing tracker power supply bank will be entirely replaced. Based on estimates presented above, the supply system will need to provide approximately 100 kW of FE power, including losses on the supply cables, at a voltage of 13-14 V and roughly equally split between 2S and PS systems. The HV system will need to generate up to  $-800$  V of sensor bias voltage, with close to 2 mA of leakage current per module at end of life in the most exposed areas of the tracker. Both LV and HV distribution systems will have a module-level granularity that fits well the autonomous module building-block philosophy and avoids implementing complicated failure mitigation schemes. Thanks to the “elevated” low voltage supply rail, ohmic losses in the supply cable will be kept at a manageable level, with an estimated voltage drop of only 3 V along 80 m. This opens the possibility to locate the power supply (PS) units in the Service Cavern, alleviating the need to qualify radiation and magnetic field tolerant units, and easing the long-term maintenance scenario. However, the option of installing radiation tolerant supplies in the experimental cavern is kept open as constraints linked to counting room floor space and cabling channels’ availability may complicate the situation.

Multi-fibre cables (comprising 144 fibres each) and multi-service power cables (comprising 12 LV and 12 HV channels each) connect on-detector to off-detector electronics. A segmentation of the Outer Tracker into eight octants per  $z$  side (i.e. 45 degree  $\varphi$ -sectors) results in a tentative rack layout with one DTC-rack per octant in the counting room (i.e. a total of 16 racks). Each rack houses two ATCA shelves, one for the 2S system (nine DTC blades per shelf) and one for the PS system (seven DTC blades per shelf). Each DTC blade receives one multi-fibre cable and thus processes the data from up to 72 modules. With only two shelves per rack, the remaining bottom rack space can be used for optical patch panels, shelf power supplies (AC-DC



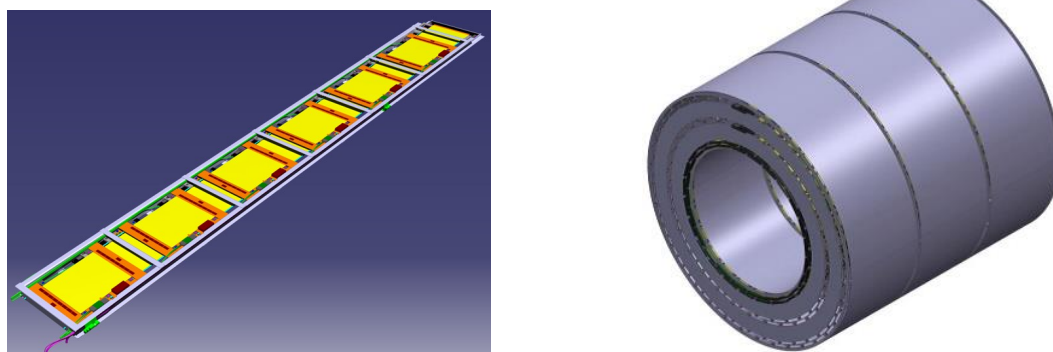


Figure 4.13: Drawings of a TB2S ladder with its 12 modules (left) and of the support wheel (right).

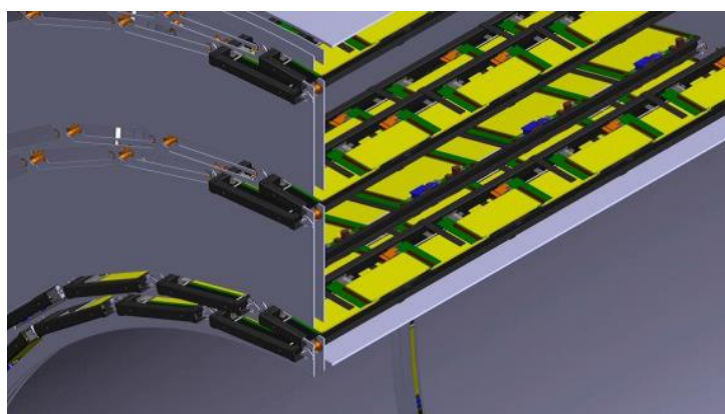


Figure 4.14: Sketch of ladders installed in the support wheel, looking at the wheel from its end.

converters) and other auxiliary equipment.

The Data, Trigger and Control board, DTC, will be heavily loaded with optical in- and outputs. A total of 72 fibre pairs will link each DTC to its associated FE modules (i.e. up to 720 Gb/s aggregate input data rate and 180 Gb/s output rate), while additional links will connect it to the DAQ and L1 Track finding systems. An aggregate bandwidth of 100 Gb/s is estimated for DAQ data and 600 Gb/s for TRIG data.

#### 4.4.4 Outer tracker mechanical structure

##### 4.4.4.1 The Tracker Barrel with 2S modules (TB2S)

The TB2S mechanics reuses proven solutions from the current tracker, adapted to the new layout and to the new of modules and electronics. The TB2S is equipped with 4464 detector modules of type 2S, all featuring a sensor spacing of 1.8 mm. The modules are mounted onto 372 ladders, each containing 12 modules (Fig. 4.13 (left)). The ladders are supported by a wheel that consists of four vertical disks joined by cylinders at the inner and outer radii (Fig. 4.13 (right) and Fig. 4.14). The ladders, which are half the length of the wheel, are installed from the two ends of the barrel and have a small overlap at the center to avoid gaps in the coverage.

The TB2S ladders are made of two parallel carbon fibre C-shaped profiles, joined by several

orthogonal elements (referred to as cross pieces), also made of carbon fibre. The modules are supported by machined inserts, made in aluminium carbon fibre composite material and supported themselves by the C-shaped profiles. These inserts provide also a connection to the cooling pipe that transits the full length of the ladder and back, forming a U-shaped circuit.

The ladder structure is assembled in a high-precision metallic jig. The components (carbon fibre C-profiles, cross pieces, inserts and cooling pipe) are positioned on the jig and a low viscosity epoxy adhesive is applied to the joints. The adhesive is cured at room temperature to avoid deformations caused by the difference in thermal expansion of the metallic jig and the carbon fibre components.

Low and high voltage wires as well as optical fibres are routed along the modules and inside the C-profiles to one end of each ladder. There will be in total 36 wires per ladder, three for each module (LV, HV, common ground). At the end of the ladder those wires are connected to one multi-core cable. The 24 optical fibres (two per module) are connected at the ladder end to two 12-fibre cables.

Small-size all-metal fittings are used to connect each ladder cooling pipe to the supply and return cooling pipework. Those fittings allow also temporary connections during the ladder manufacturing and testing phases.

#### 4.4.4.2 The Tracker Barrel with PS modules (TBPS)

The TBPS has 2836 PS modules distributed on three layers. Each layer is sub-divided into one central section and two end sections, as shown in Fig. 4.15. The modules in the central section are horizontal (barrel arrangement) and supported by flat plank structures. The modules in the end sections are tilted towards the interaction point with tilt angles ranging from 47 to 74 degrees. The tilted modules are supported by circular rings. Several successive rings have the same tilt angle, thus limiting the number of different ring geometries to be produced. The flat and tilted sections of a layer are joined mechanically with longitudinal I-beams. These beams provide also a support for routing the cooling pipes, LV and HV wires and optical fibres to the two ends of the TBPS. The TBPS is supported from horizontal rails attached to the inner cylinders of the TB2S wheel.

The planks of the flat central section, shown in Fig. 4.16 (left), are of sandwich construction with a foam core and two carbon fibre skins. The cooling pipe is in the mid-plane of the plank. It transits the full length of the plank and returns back to make a U-shape. The pipe is embedded into carbon foam, which acts as a heat spreader and thus improves the thermal connection to the cooling pipe. Lighter foam (Airex) is used where thermal connections are not needed, i.e. at the end of the plank and inside of the U-shaped cooling circuit. Small metallic inserts, embedded into the plank structure, allow the positioning of the modules and attaching of the plank to end support rings (Fig. 4.16). Phase-change thermal interface material is used to make the thermal connection between the modules and the plank surface. These connections can be dismantled by warming up to about +50°C temperature, when the thermal interface material becomes soft.

On each plank the wires and optical fibres are routed to one end of the plank and then further towards the end of the TBPS. The cooling pipe connections are also at that same plank end. In each layer half of the planks have their services routed to the +z end, the other half to the -z end.

The modules on the tilted TBPS end sections are placed on ring structures made of carbon fibre / foam sandwiches. One tilted ring and one of the layer 1 tilted sections are shown in

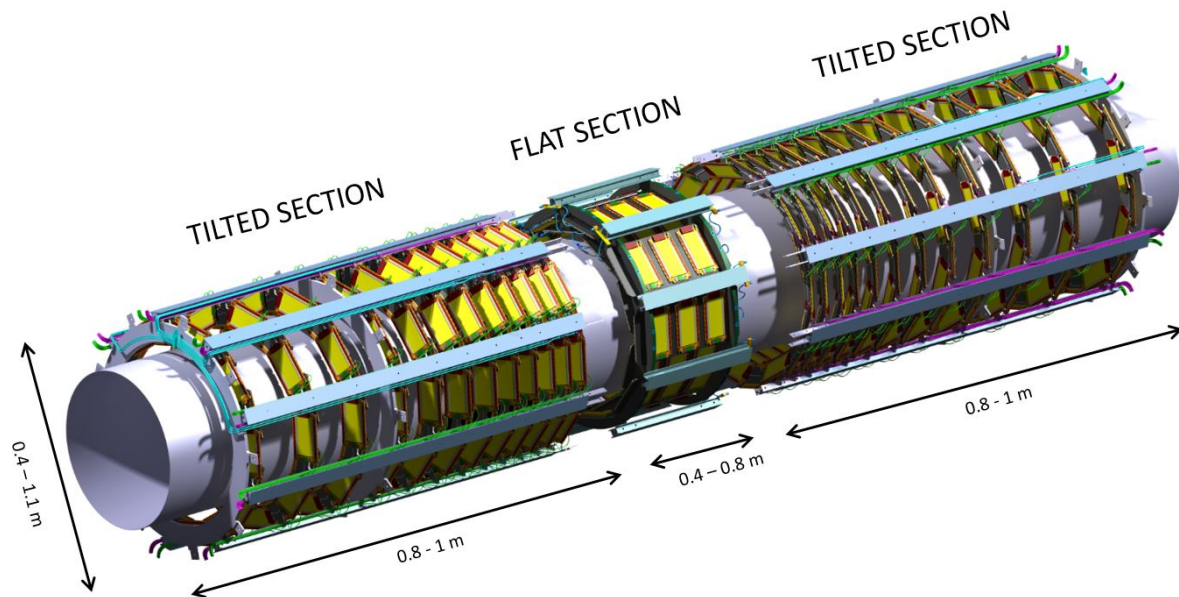


Figure 4.15: Drawing of the innermost layer (layer 1) of the TBPS, showing the central flat section and the two (identical) tilted sections. The length of each section varies between the three TBPS layers. The layer 1 also integrates the central part of the Inner Tracker support tube, as visible inside.

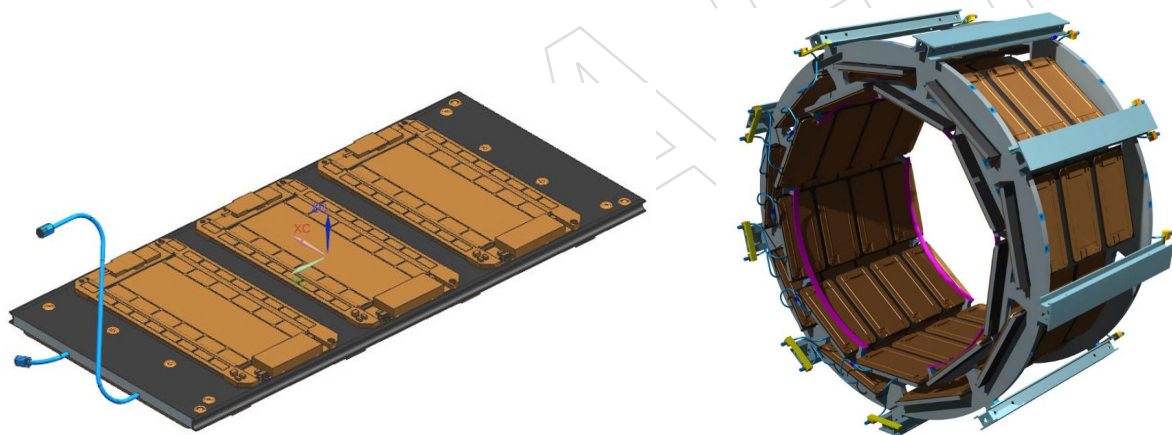


Figure 4.16: A TBPS layer 1 plank and the layer 1 central flat section with its 18 planks.

2439 Fig. 4.17. Half of the modules are placed on the front side of the ring, half on the back side.  
 2440 Correspondingly there is one circular cooling pipe on each side of the ring. Cooling plates  
 2441 in high-conductivity carbon fibre / epoxy laminate are used for attaching the modules to the  
 2442 support ring and to make the thermal connection. Machined pieces in aluminium / carbon  
 2443 fibre composites make the connections between the cooling plates and the cooling pipe. Like in  
 2444 the flat TBPS planks a phase-change thermal interface material is used to provide the thermal  
 2445 contact to the modules.

2446 Assembly of the TBPS rings is done on dedicated precision jigs, using room-temperature cured  
 2447 low-viscosity epoxy adhesive. For each of the eight different types of rings used in the TBPS  
 2448 two jigs are needed, one for the front side and a second for the back side.



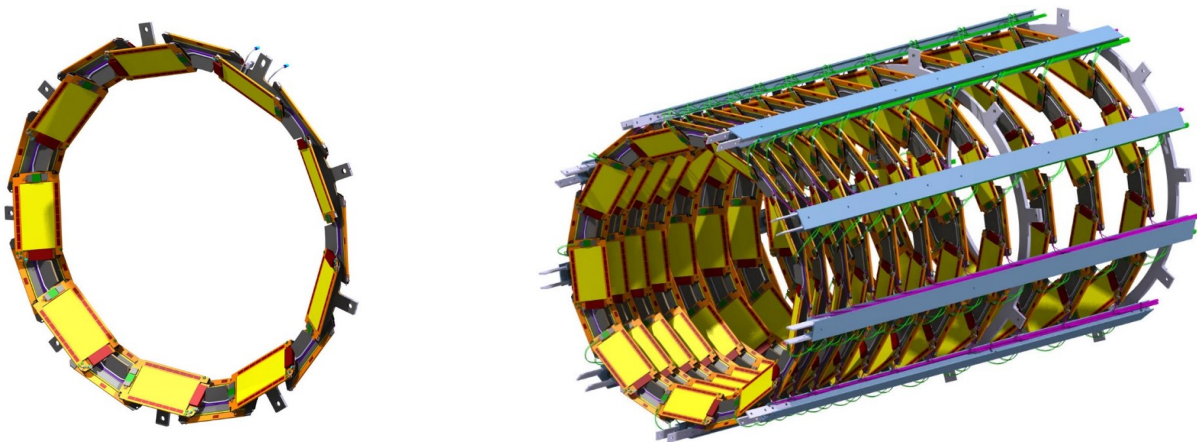


Figure 4.17: A TBPS layer 1 tilted ring and one of the two layer 1 tilted sections.

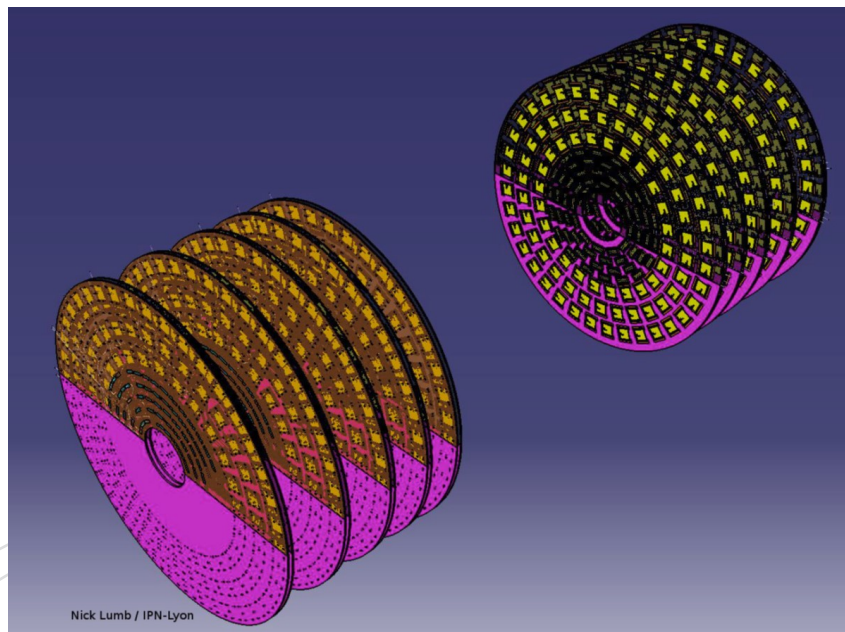


Figure 4.18: The two identical TEDD units, each consisting of five double-discs. Each double-disc consists of four dees. **FIXME:** Get picture with modules everywhere? without pink?

#### 4.4.4.3 The Tracker Endcap Double-Discs (TEDD)

The TEDD uses exactly the same rectangular 2S and PS modules that are used in the barrel sections of the Outer Tracker. The modules are mounted on flat discs, which for assembly reasons are split in half-discs, or “dees”. Two discs are grouped to form one double-disc, which provides one hermetic detector plane. Ten double-disc units will be produced, five for each endcap (Fig. 4.18). To accommodate the change in diameter of the Inner Tracker support tube, the TEDD disks need to be of two different inner radii.

The cooling pipes run inside the sandwich structure of each dee. There are seven cooling circuits in each dee, organised in sectors (Fig. 4.19 (left)). Carbon foam blocks are used in the PS module part of the dee to provide thermal connection between the dee skins and the cooling pipe. The PS modules, located in the low radius regions, are thus attached directly to the cooled surfaces of the dee. Aluminium inserts are used to provide thermal contacts between the 2S

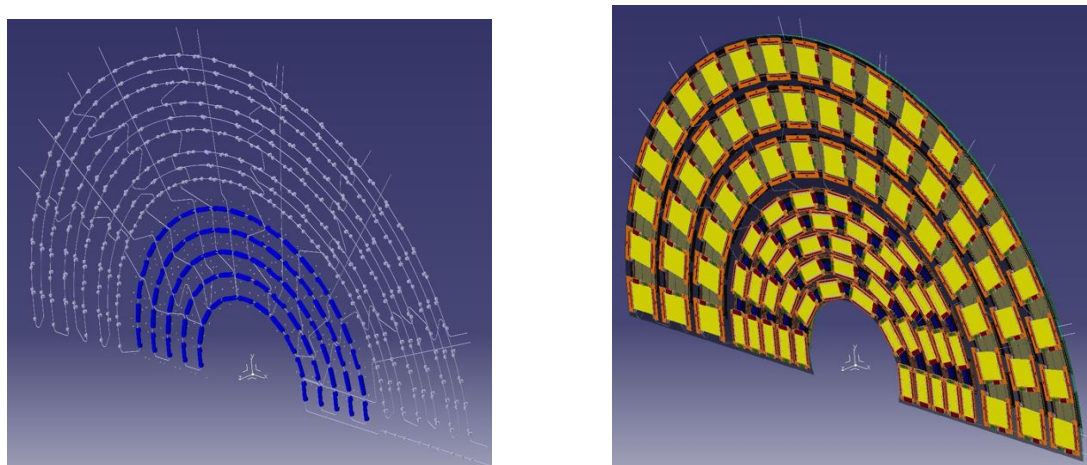


Figure 4.19: The seven cooling circuits of one dee (left), and a fully assembled dee (right).

modules, located in larger radius regions, and the cooling pipes. Where no thermal connection is needed a lighter and cheaper polyetherimide foam (e.g. Airex R82) is used. To obtain a flat dee the different foam sections are machined to even levels. The module support inserts are guided to their positions using a precision assembly table. After the gluing of the dee the inserts both on the top and on the bottom side are machined to obtain the final module support and guide surfaces.

Power wires and optical fibres are routed radially outwards, above and between the modules, to the outer periphery of each dee. Thin protection frames are used to constrain the services and protect the modules. In the complete TEDD assembly of five double-discs the services run to the end of the TEDD and have there sufficiently long pigtails to reach the tracker patch panels (PP1) in CMS.

Each TEDD unit is installed to the tracker support tube as one package. The two TEDD units are identical.

Details about common mechanics and services can be found in Appendix .1.

#### 4.4.5 Outer tracker prototyping, system tests, and beam tests

During the development phase of the silicon sensor modules and sub-assemblies prototypes are being built and evaluated in view of performance and robustness. System tests validate the performance of the modules in operational conditions as close as possible to the expected running conditions in the experiment. This includes the study of potential negative influence of environmental conditions (temperature, humidity) outside the nominal expectations on the performance, as well as potential electromagnetic interference effects. Beam test experiments are conducted to study the behaviour of the modules when subjected to a particle beam, and in particular to verify the performance of the stub finding mechanism.

Each Outer Tracker module constitutes a self-contained and independent unit, carrying its own components for power delivery (DC-DC converters) and opto-electrical conversion on the service hybrid. As a consequence a single module can be considered already a “system”, while larger sub-structures combine several of these identical systems on one mechanical support structure.

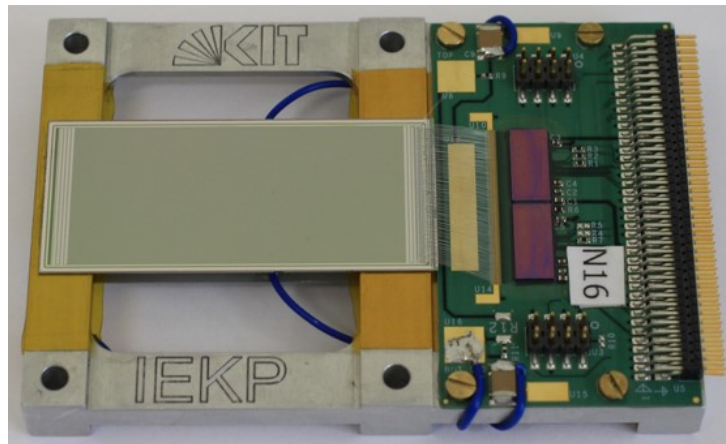


Figure 4.20: A 2S mini-module assembled from a small prototype hybrid comprising two CBC2 readout chips and two 5 cm long strip sensors, mounted on top of each other in an aluminium frame.

#### 4.4.5.1 Module prototyping

Different types of prototype modules were built and used for system tests and test beam measurements. In the following the different 2S module prototypes are described. The PS-prototypes were built in the US and are documented in Sec. 4.5.5.

- **2S mini module:** This module type, shown in Fig. 4.20, consists of a small double-sided rigid hybrid with two CBC2 readout chips and two small strip sensors arranged on top of each other (as in the full 2S modules), each with 254 strips of 5 cm length. The spacing of the two sensors is realized with a simple frame made of aluminium. In total about ten mini-modules were built by two institutions, and used for a variety of tests. In particular, this module prototype has been studied in several beam tests.
- **2S full-size module:** This early version of the 2S module comprises two full-size 1024 strips 2S sensors. The sensors are connected to two folded Kapton hybrids containing eight CBC2s each, but no CIC. Thus direct connections to each CBC2 are routed out to a connector. Low voltage as well as high voltage are provided externally, as the service hybrid is not integrated. Three complete prototype 2S modules with 1.8 mm spacing have been produced at CERN with working sensors and front-end hybrids. One of these modules is shown in Fig. 4.21. All three modules were successfully constructed and wire bonded, with all readout channels functioning correctly. Although these three modules had some parts that were not as for the current baseline design (for instance a different Kapton thickness and different glue was used between the Kapton isolator and the sensor), the sensors and FE hybrids were of the latest versions and the results from a test beam run on one module and bench tests of all three modules indicated that the modules perform as expected. Much useful experience has been gained during module prototyping, which has resulted in design improvements in both the module and the assembly jigs, as well as modification of materials, such as glues and HV isolation layers. Further prototypes are in progress to test the assembly of the 2S module version with 4 mm sensor spacing and to produce more modules with the latest baseline design, so that reliability studies can be performed in addition to production method qualifications. These modules will also be used for further performance studies of the sensors and front-end electronics



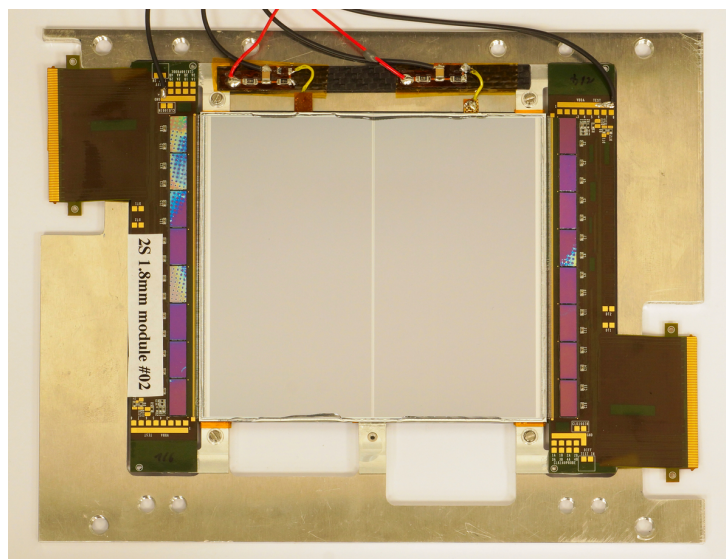


Figure 4.21: A full-size 2S module (1.8mm variant) comprising two flex hybrids with eight CBC2s each, two 10 cm long strip sensors and aluminium bridges.

(Sec. 4.5.5).

- MaPSA-light: This bare module is the bump bonded assembly of up to six small prototype MPAs (MPA-light) with 48 readout channels each to a small version of the macro-pixel sensor (PS-p sensor). Around 20 MaPSA-light assemblies were built both in-house and at four different vendors, with a focus on the optimization of the bump bonding process. The analogue functionality and hit efficiency of this prototype module has been evaluated in beam tests (Sec. 4.5.5).
- PS micro module: To evaluate the stub finding features of the MPA two MaPSA-light assemblies are stacked on top of each other with a few millimeter gap. One MPA is set to strip emulation mode and the other MPA can process those data to send out stubs. This will allow to verify the functionality of the stub formation in beam test experiments or using radioactive sources.

#### 4.4.5.2 System tests

From a system level point of view there are several aspects which need to be tested early in the prototyping phase:

- Components of the modules like sensors, service hybrids and front-end hybrids are closely spaced in the final arrangement and should not influence each other. This can be studied for single prototype modules, e.g. by comparing the performance of the components before and after assembly. An example of such a test based on a prototype of the service hybrid for 2S modules is described below. Operation of prototype modules in the foreseen arrangements on the substructures (ladder, ring, plank, dee) are the next step.
- Prototype modules shall be operated for durations in the order of weeks to check for stable performance. Special attention is given to the operation of modules at temperatures similar to those expected to be reached in the detector. Such tests have successfully been conducted with a mini-module, taking the stability of noise versus time as a figure of merit, and are in preparation for full-size modules, which requires

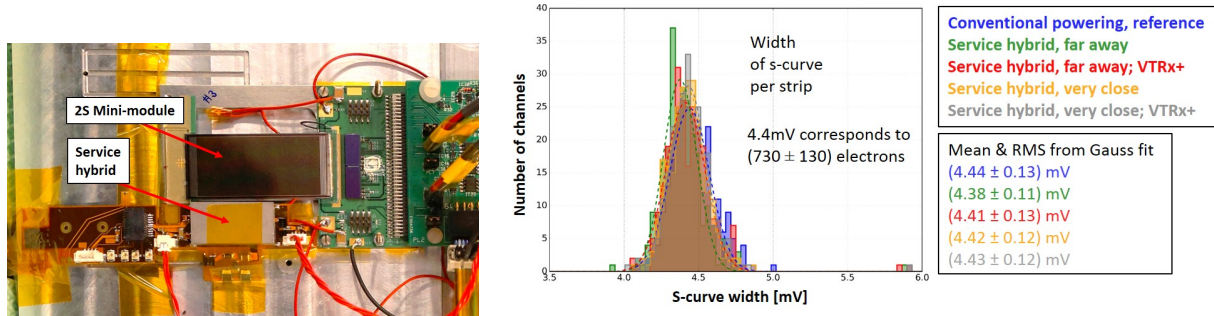


Figure 4.22: Photo of the system test setup with a 2S mini-module (left), and histograms of the noise of all strips of one CBC2, measured with conventional powering (blue), a SH placed far away (green) and as close as shown in the photo (yellow), plus measurements with data flowing through the VTRx+ (red and grey).

a sufficiently sized cold box due to the necessity to arrange large readout adapter cards.

- A central and new feature of the modules is the ability to discriminate track momenta by their curvature in the magnetic field of CMS. This has been evaluated in beam test experiments with early prototypes of 2S mini-modules and full-size 2S modules.
- The front-end electronics provides sparsified stub data for each bunch crossing (at 40 MHz) plus binary data (unsparsified for 2S modules and sparsified for PS modules) after level-1 accept (at up to 750 kHz). The readout chain needs to sustain the envisaged data rates and the limit on occupancy, i.e. the occupancy at which data starts to get lost, need to be tested. The entire chain will undergo intensive tests with injected charge pulses or in an appropriate test beam environment.

In the most complete system test performed so far, a mini-module was combined with a prototype of the service hybrid (Fig. 4.22, left). The module received its low voltages from the service hybrid, which was connected to a prototype of the low voltage power supply via a cable of realistic length. The module's noise was extracted and a comparison made between the noise obtained when the module was powered from a laboratory power supply and the noise present when the module was powered via the service hybrid and consequently two DC-DC converters (Fig. 4.22, right). No significant difference in noise was observed, indicating that the operation of DC-DC converters in close vicinity to the sensors and readout hybrids is feasible and does not deteriorate the module's performance. Similar measurements with full-size modules are ongoing at the time of writing.

#### 4.4.5.3 Beam tests

The performance of prototype and full-size modules has been assessed by operating them in test beam experiments at DESY, CERN, and FNAL. The test beam campaigns at DESY and CERN were dedicated to the 2S assemblies, while the PS prototypes were studied at the FNAL test beam facility. The data collected during those beam tests are utilized to measure the cluster reconstruction efficiency and cluster width, and the stub reconstruction efficiency as a function of the beam incidence angle. The latter allows to estimate the stub reconstruction efficiency as a function of the track transverse momentum.

Figure 4.23 presents the stub efficiency of a non-irradiated 2S mini-module and that of a device irradiated to a fluence of  $6 \times 10^{14} \text{ n}_{\text{eq}}/\text{cm}^2$ . The measurement was carried out at the CERN

H6B beam line using 120 GeV pions. Tracks are reconstructed using data from the AIDA telescope [68] and matched to hits in trigger planes, which utilize the ATLAS FE-I4 chip [69]. The stub reconstruction efficiency is defined as the ratio of the number of events with at least one track-matched stub to the number of events with one track matched to the trigger plane hits. Tracks and stubs have to match within  $4\sigma$  of the spatial resolution. Figure 4.23 (top) shows that the response of the detector is uniform throughout the sensor's region where the beam was incident. Figure 4.23 (bottom) compares the performance of the irradiated and non-irradiated device. For the non-irradiated module the measured turn-on threshold, defined as the  $p_T$  for which the stub efficiency reaches 50%, is 1.88 GeV with a  $p_T$  resolution of 5%<sup>4</sup>, to be compared to an expected turn-on threshold value of 2 GeV. The sharp turn-on and the plateau value of 99% efficiency demonstrate that this module type can select efficiently stubs above the chosen threshold and thus meets the specifications. The plateau efficiency of the irradiated mini-module remains above 95% and the  $p_T$  resolution is preserved at 5%. The turn-on thresholds in the non-irradiated and in the irradiated mini-module differ due to a different sensor spacing in the two devices; no attempt was made to tune the stub windows such that exactly the same turn-on threshold is achieved.

Results from a beam test carried out at the Fermilab Beam test facilities on PS-prototypes built in the US is documented in Sec. 4.5.3 as well as the characterization of the one of the CERN-built full size modules.

#### 4.4.6 Outer tracker expected physics performance

This section describes the expected physics performance of the Phase 2 tracker with focus on the Outer Tracker.

##### 4.4.6.1 Material budget

The material description uses information from engineering models where available and it is based on components from the current tracker otherwise. A list of materials with average densities and characteristic radiation and interaction lengths was compiled and these materials were assigned to the supporting structures, the silicon modules and their services. The materials and assignments served as input to the TkLayout software [70, 71], a dedicated tool developed in the context of the tracker Phase-2 Upgrade project in order to study the effect of design choices on the overall layout.

The routing of the services is automatically performed by tkLayout (Fig. 4.24) and the grouping

---

<sup>4</sup>The correlation logic of the CBC2 has been tested by rotating the device, thus emulating the bending of tracks in the magnetic field. The beam incident angle,  $\alpha$ , represents a particle with a certain bending radius in the transverse plane,  $r_T$ , for a given radial position of the module in the tracker,  $R$ , following the equation  $\sin(\alpha) = \frac{R}{2 \cdot r_T}$ . This bending radius corresponds to a particle with certain charge,  $q$ , and transverse momentum,  $p_T$ , for a homogeneous magnetic field of given strength,  $B$ , via the relation  $r_T = \frac{p_T}{q \cdot B}$ . For typical units, under the assumption of  $q = \pm e$ , this reads as

$$r_T[\text{m}] \sim \frac{p_T[\text{GeV}]}{0.3 \times B[\text{T}]} \quad (4.1)$$

For the CMS field strength of  $B = 3.8 \text{ T}$ , the relationship between the beam incident angle and the transverse momentum of the traversing particle for a radial position of the module is given by

$$p_T[\text{GeV}] \sim \frac{0.57 \times R[\text{m}]}{\sin(\alpha)} \quad (4.2)$$

Data are fitted using an Error Function and the  $p_T$  resolution is calculated as the ratio of the width of the Gaussian to the  $p_T$  value at 50% of the plateau height.

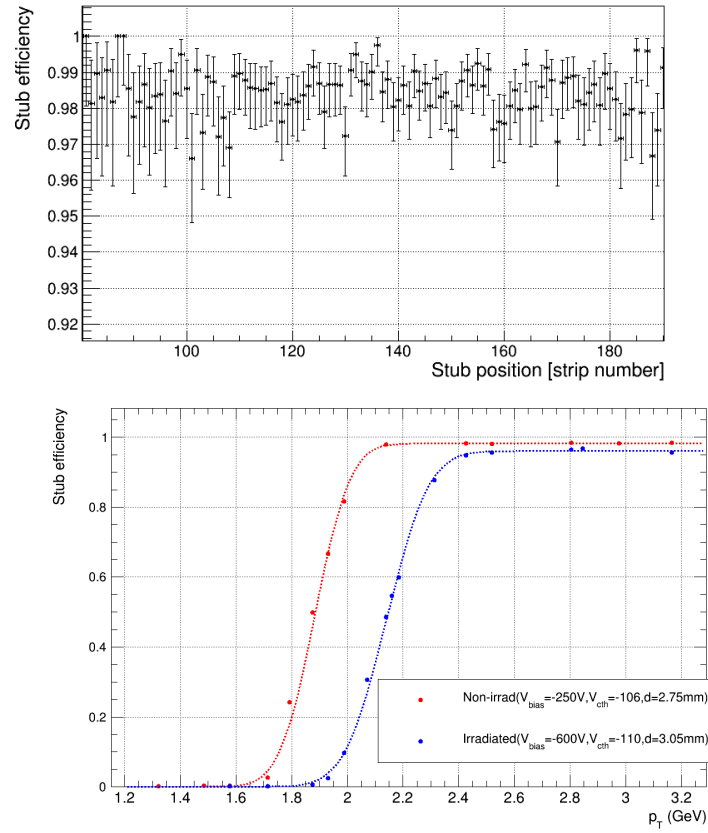


Figure 4.23: Top: stub reconstruction efficiency of the unirradiated 2S mini-module presented as a function of the stub position. Strip numbers 0 to 126 correspond to the first CBC, strip numbers 127 to 253 belong to the second CBC. Only the region where beam was incident is displayed. The error bars correspond to statistical uncertainties. Bottom: stub reconstruction efficiency for a non-irradiated (red) and an irradiated (blue) 2S mini-module. The mini-module was irradiated to a fluence of  $6 \times 10^{14} \text{ n}_{\text{eq}}/\text{cm}^2$ .

of the services (e.g. the distribution of cooling fluid at the cooling manifolds) is also defined within that software framework.

Outer Tracker modules are described by eight possible categories (referred to as “volumes”) representing the three or four hybrids, the two sensors, a support plate below the sensors in PS modules, and the spacers between the two sensors. Inner Tracker modules are described by three volumes, representing the sensor, the material above it, and the material below it. Each module material category is assigned to the appropriate volume in the three-dimensional description.

All services are assigned to cylindrical volumes running close to the barrel layers and to the discs at higher radii and  $z$  positions, respectively, with a variable density taking into account the accumulation of services along the structures, where appropriate.

The impact of the detector’s material on tracking resolution and secondary interactions is best evaluated by comparing the amount of material inside the *tracking volume*, defined as the material crossed by a straight line between the origin and the farthest silicon sensor along the line. Figure 4.25 shows a comparison of the estimates of the material inside the tracking volume, in

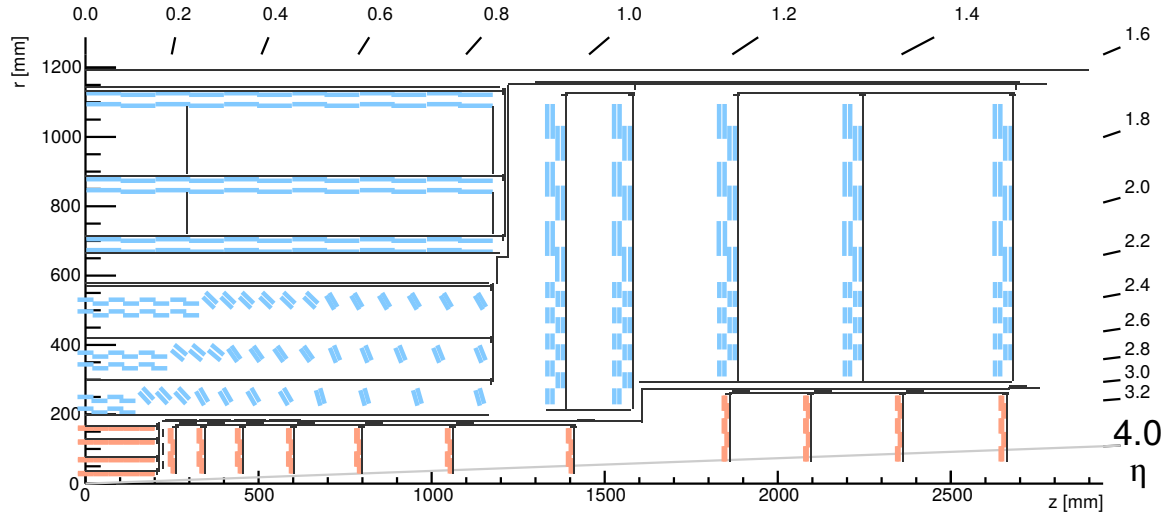


Figure 4.24: Position of services and modules in the material budget model, in cylindrical coordinates and summing over  $\varphi$ . Service volumes are represented by black lines. For each module the average sensor position is represented by a single line. Blue lines represent Outer Tracker modules and red lines Inner Tracker modules.

units of radiation lengths, for the Phase-1 detector (as currently defined in the standard CMS simulation and reconstruction software CMSSW [72]) and the Phase-2 detector (as estimated with tkLayout). The material budget of the Phase-2 detector is slightly smaller than that of the Phase-1 detector in the centre of the detector, i.e. at  $\eta \approx 1$ , and significantly smaller in the region around  $\eta = 1.5$ .

#### 4.4.6.2 The simulation setup

A full Monte Carlo simulation of LHC events with high pileup gives the best prediction of the detector performance and overall physics capabilities of the upgraded CMS detector, once the detector response is known and correctly simulated (including sensing elements, electronics and inactive material), and the reconstruction algorithms have been fully optimized for the new detector.

For the studies presented in this document, the CMS detector response was simulated using the official CMS software package CMSSW, also used for the simulation of the present detector and for the analysis of data collected so far. The detector geometry and materials are described and simulated using the Geant4 toolkit. The simulation also includes and uses information about the magnetic field. CMSSW reads the individual generated events and simulates the effects of energy loss, multiple scattering and showering in the detector materials with Geant4. The digitization (simulation of the electronic response), the emulation of the level-1 trigger, and the offline reconstruction of physics objects are performed as well.

In the simulation, the thickness of all sensors in the Outer Tracker is set to  $200 \mu\text{m}$ , while  $150 \mu\text{m}$  thick sensors are simulated for the Inner Tracker. The simulated pixel size amounts to  $25 \times 100 \mu\text{m}^2$ . The strips in 2S and PS-s sensors have simulated dimensions of  $90 \mu\text{m} \times 50.25 \text{ mm}$  and  $100 \mu\text{m} \times 23.13 \text{ mm}$ , respectively. Finally the macro-pixels in PS-p sensors were simulated with a pitch of  $100 \mu\text{m}$  and a length of  $1446 \mu\text{m}$ .

The geometry implementation is automatically derived from the detailed representation developed within the tkLayout tool discussed in Sect. 4.4.6.1. In addition to the sensitive volumes



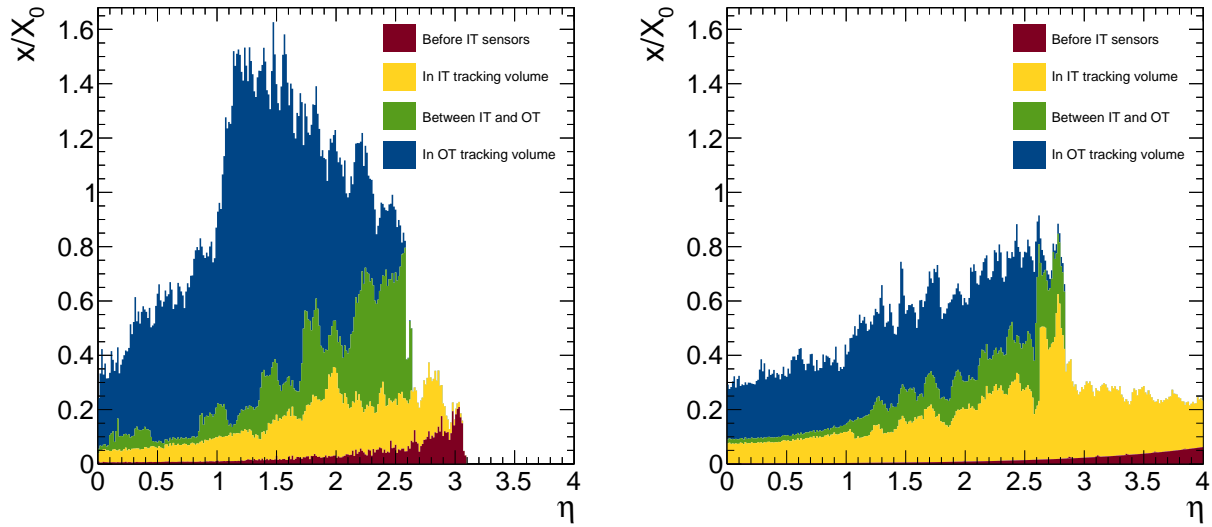


Figure 4.25: Material budget inside the tracking volume estimated in units of radiation lengths, comparing the Phase-1 detector (left) with the Phase-2 detector (right). The material in front of the Inner Tracker sensors is shown in brown, that inside the Inner Tracker tracking volume is shown in yellow, the material between IT and OT sensors is shown in green and the material inside the Outer Tracker tracking volume is shown in blue. The histograms are stacked. **FIXME:** Improve label.

the support structures and services are included, with a submillimetric level of detail.

#### 4.4.6.3 Hit simulation and performance

The simulation of the readout electronics of the detector modules used to acquire data by the tracker takes place in the so-called digitization step. A common simulation package has been developed for both Inner Tracker and Outer Tracker detectors. It starts from the hit positions and simulated energy losses in the sensors and produces an output that is as close as possible to the real data coming from the detector. The energy loss of a charged particle crossing a silicon layer is distributed along a path between the entry and exit points within the silicon sensor, taking into account the Landau fluctuations. The charge produced in the sensor is migrated to the sensor surface properly taking into account the Lorentz drift and the diffusion in the perpendicular plane, and assigned to the detector channel (pixel, macro-pixel or strip) at that position.

The Gaussian noise from the readout electronics is added on top of the collected charge for each channel. Noise is also added to other channels taking into account crosstalk, and may generate extra hits. Channels with exceptionally large noise or inefficiency can also be simulated at this step.

Following the integration of all these contributions, a digitized hit (“Digi”) is created when the total charge associated with a given channel exceeds a predefined threshold. The optimization of the threshold and readout parameters has been done separately for each type of sensor and readout electronics. The threshold is 0.4 times the most probable charge from a minimum ionizing particle (MIP). This guarantees high efficiency while maintaining an almost negligible noise contribution. In the Short Strip ASICs (SSAs) of PS modules an extra bit is set when the charge exceeds 1.4 MIPs, to help with the identification of highly ionizing particles. Signals from particles crossing silicon sensors within  $\pm 12.5$  ns from the nominal bunch crossing are considered.



This includes particles from that bunch crossing but also looping particles from up to five previous crossings. For 2S modules the parameters are tuned using beam test results obtained between 2013 and 2016, while for PS modules the parameters are set according to the present understanding of the readout electronics and will be tuned in future with corresponding beam tests.

The occupancy obtained with this setup, defined as the fraction of active channels, is presented in Fig. 4.26 for events with 200 pileup events. Even at this high level of pileup, the occupancy remains at an acceptable level, not exceeding 3% in the first TBPS layer. The hierarchy visible in the inner layers is also a driver for the seeding strategy used for tracking. The reduction in occupancy resulting from the tilted design is clearly visible around  $|\eta| = 0.5$  for the three TBPS layers.

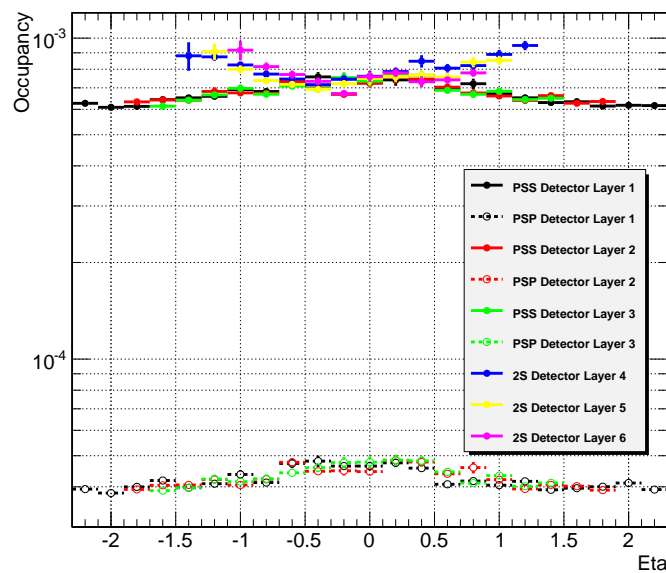


Figure 4.26: Hit occupancy, defined as the fraction of channels containing a digitized hit, as a function of  $\eta$  for all layers of the TBPS and TB2S. **FIXME:** Plot is still for zero pileup. Should include endcaps as well. Improve legend and labels.

In the Inner Tracker, clusters are built from connected pixels. The position of the cluster is obtained as the barycenter of the cluster. In the Outer Tracker, clusters are built in the high-resolution direction only, by aggregating adjacent hits. The position of the cluster is taken as its geometric centre. This cluster position is corrected for the Lorentz drift and a position uncertainty is estimated.

#### 4.4.6.4 Stub simulation and performance

In the simulation stubs are obtained from dedicated clusters built independently from the clusters used in the offline reconstruction. In fact clusters are built and combined into stubs following the exact algorithms implemented in the CBC and MPA chips. The studies presented in this section use samples of simulated muons.

Compared to the flat version of the tracker barrel geometry considered previously in the CMS Phase-2 Technical Proposal, the tilted geometry allows to maintain high stub reconstruction efficiency over the full OT acceptance. Figure 4.27 shows the stub efficiency in the innermost

2701 TBPS layer, comparing both geometries and simulating the readout system **FIXME**: Should  
 2702 mention the  $p_T$  threshold. An efficiency above 90% is obtained over the full range for the tilted  
 2703 geometry, while the efficiency decreases to 40% at large pseudorapidities for the flat geometry.  
 2704 This difference is due to the lack of communication between the two sides of PS modules,  
 2705 as each row of strips is read out by one separate front-end hybrid. The tilted geometry still  
 2706 provides a good transverse momentum discrimination in the stub reconstruction, as illustrated  
 2707 in Fig. 4.28 for simulated muons. The turn-on curves depicted in that figure are sharp, with an  
 2708 efficiency compatible with zero below 1 GeV and reaching the maximal value between 2 GeV  
 2709 and 4 GeV. The turn-on is sharper at larger radii, where the stub  $p_T$  resolution is better.

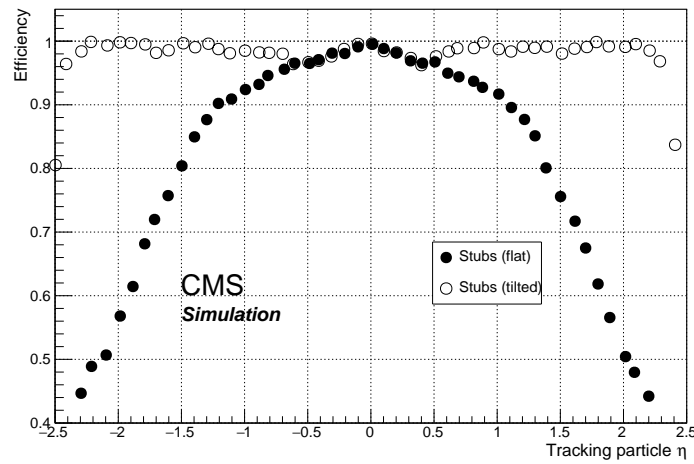


Figure 4.27: Stub reconstruction efficiency in TBPS layer 1 as a function of  $\eta$ , comparing the flat (solid points) and tilted (open circles) tracker barrel geometries. **FIXME**: Improve legend and labels, add  $p_T$  threshold.

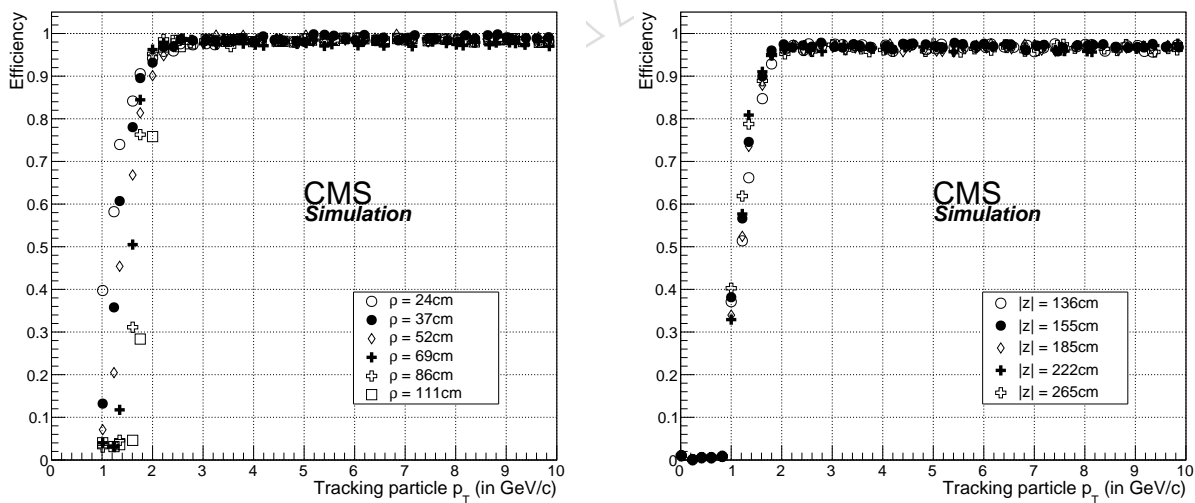


Figure 4.28: Stub reconstruction efficiency for muons as a function of  $p_T$  in the barrel (left) and in the endcap regions (right). **FIXME**: Improve legend and labels.

#### 4.4.6.5 Offline tracking performance

**FIXME:** All offline tracking plots will be updated once the final TDR tracker geometry is available in CMSSW.

In each bunch crossing, the CMS tracker will be traversed by around 4000 tracks for average pileup values of 140. In this section preliminary results on the offline tracking performance over the full acceptance of the CMS tracker are presented.

The performance of the track reconstruction can be summarized by the track finding efficiency, the fake rate, and the resolution of the estimated track parameters. Two simulated samples have been used: single muons with a transverse momentum of 10 GeV and  $t\bar{t}$  events, both with superimposed pileup of minimum-bias events. Two pileup scenarios – denoted 140PU and 200PU – are considered, where the number of pileup events was drawn from a Poisson distribution with mean equal to 140 or 200, respectively.

Figure 4.29 shows the tracking efficiency for single muons in both pileup scenarios. The single muon efficiency is stable and close to 100% in the entire range of pseudorapidity and does not change with the amount of pileup.

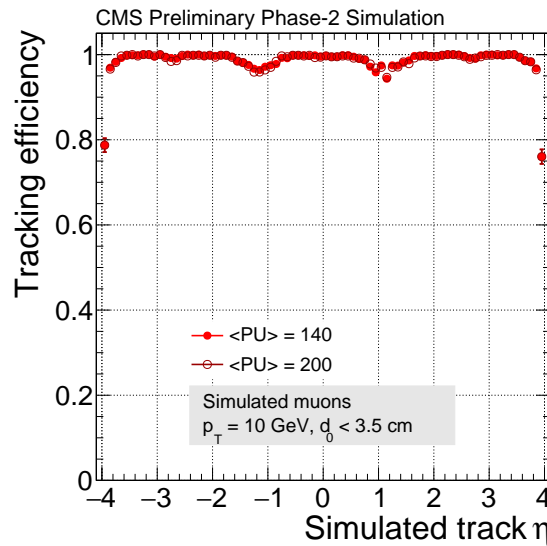


Figure 4.29: Tracking efficiency as a function of the pseudorapidity for single muons with  $p_T$  equal to 10 GeV, with 140 pileup events (full circles) and 200 pileup events (open circles). The efficiency is shown for tracks produced less than 3.5 cm from the centre of the beam spot in the radial direction.

Figure 4.30 shows the results for  $t\bar{t}$  events in both pileup scenarios. The efficiency and the fake rate for tracks with  $p_T > 0.9$  GeV are shown as a function of the pseudorapidity,  $\eta$ . The histograms contain only tracks passing a certain set of quality requirements (referred to as *High Purity* requirements). Only tracks produced less than 3.5 cm from the centre of the beam spot in the radial direction are used for the efficiency calculation. The efficiency is around 90% in the central region, dropping off at  $|\eta| > 3.8$ , while the fake rate is lower than 10% in the entire range of  $\eta$  for 140 pileup events.

A trajectory can be fully described by five parameters at the impact point, which is the point of closest approach of the track to the assumed beam axis. In the CMS convention these track parameters are:  $d_0$ , the transverse impact parameter;  $z_0$ , the longitudinal impact parameter;  $\phi$ ,

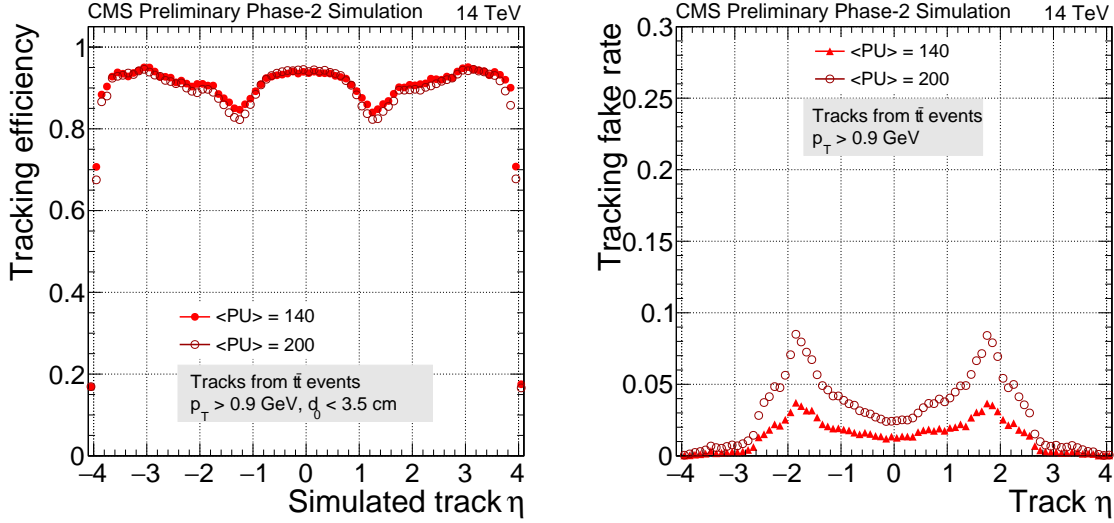


Figure 4.30: Tracking efficiency (left) and fake rate (right) as a function of the pseudorapidity for  $t\bar{t}$  events with 140 pileup events (full circles) and 200 pileup events (open circles). The tracks are required to have  $p_T > 0.9 \text{ GeV}$ . The efficiency is shown for tracks produced less than 3.5 cm from the centre of the beam spot in the radial direction.

the track angle in the transverse plane;  $\cot \theta$ , the cotangent of the polar angle; and  $p_T$ , the transverse momentum. The resolution of track parameters is given by the RMS of the residuals, i.e. the differences between the estimated and the simulated track parameters. Figure 4.31 shows the resolution of the transverse momentum and the transverse impact parameter for single muons with  $p_T = 10 \text{ GeV}$  as a function of the pseudorapidity for both the current detector and the future Phase-2 tracker.

The  $p_T$  resolution deteriorates for large  $\eta$  because of the shorter lever arm in the projection to the bending plane. Still, the better hit resolution of the Phase-2 tracker and the reduction of the material budget results in a significantly improved  $p_T$  resolution, as shown in the figure. The transverse impact parameter resolution is also improved with respect to the Phase-1 detector, ranging from below  $10 \mu\text{m}$  in the central region to about  $20 \mu\text{m}$  at the edge of the acceptance.

The CMS vertex reconstruction [73] is a two-step procedure consisting of vertex finding and vertex fitting. The primary vertex is identified as the vertex with the largest  $\sum p_T^2$ .

The key performance figure is the probability of reconstructing the signal primary vertex and to tag it correctly as the signal vertex. This is strongly correlated with the amount of pileup in the event. For  $t\bar{t}$  signal events, the efficiency to identify the primary vertex correctly is 94% with 140 pileup events, and 89% with 200 pileup events. The algorithm is not yet optimized for vertex reconstruction at very high pileup.

Figure 4.32 shows the resolution of the vertex position in the  $x$ ,  $y$  and  $z$  coordinates as a function of the number of tracks associated to the vertex.

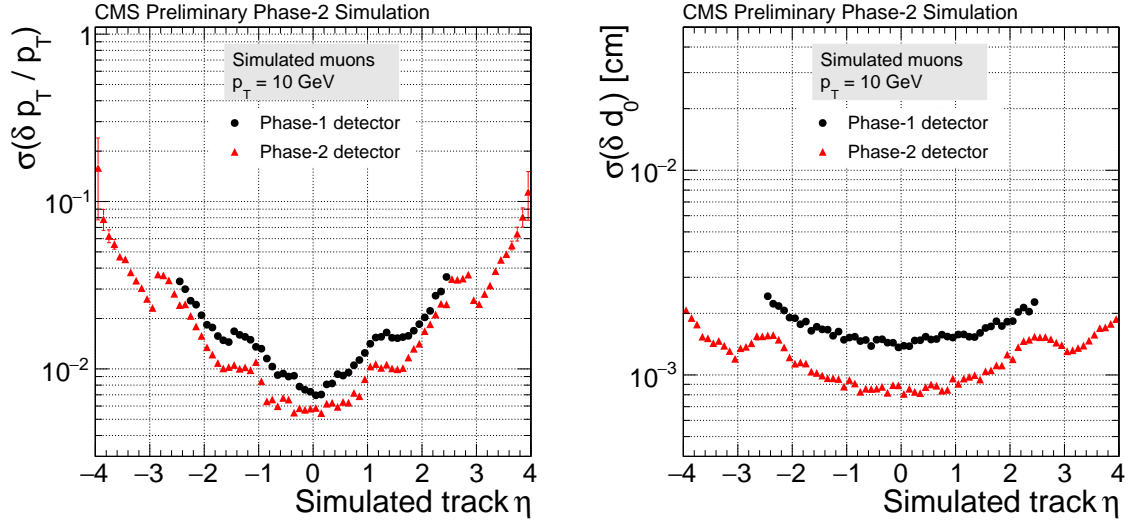


Figure 4.31: Relative resolution of the transverse momentum (left) and resolution of the transverse impact parameter as a function of the pseudorapidity for the Phase-1 (black dots) and the upgraded (red triangles) tracker, using single isolated muons with a transverse momentum of 10 GeV.

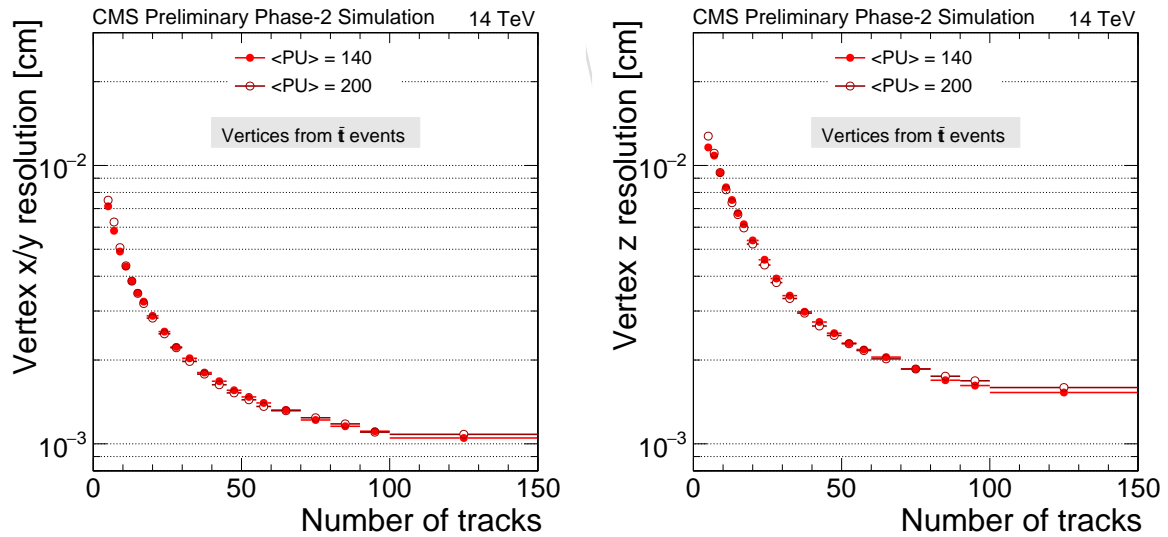


Figure 4.32: Vertex position resolution in  $x$  and  $y$  (left) and  $z$  (right) as a function of the number of tracks associated to the vertex, for  $t\bar{t}$  events with 140 pileup events (full circles) and 200 pileup events (open circles).

## 4.5 US CMS contributions

The following sections will describe in details the R&D activities carried out by US CMS within the context of the international project. Each section will also contain a brief summary of what the US CMS deliverables are for a given area (sensor, modules, electronics, mechanics, functional tests). In order to streamline the text, the general description of the various components is only presented in the preceding sections (Sec. 4.3 to Sec 4.4.6).

### 4.5.1 Sensors

this section is not final yet. We just include an image Fig. 4.5.5 for later

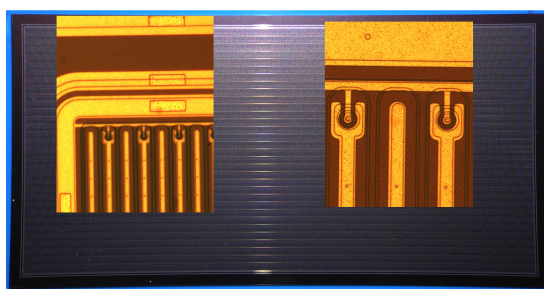


Figure 4.33: Image of a prototype PS-p sensor produced by Novati. Insets show the corner region and macro-pixel punchthrough bias structures and p-stops

### 4.5.2 Modules

The US CMS commitment is to build a total of 3450 PS and 2300 2S modules, which includes 15% spares. The first 150 PS and 100 2S modules are considered “pre-production” modules. Modules will be built by two US CMS assembly sites: Fermilab and the East Coast Assembly Site consisting of Brown, Rutgers, and Princeton. The two US CMS assembly sites will receive and inspect the necessary parts, assemble the modules, and test completed modules prior to their installation on planks and/or prior to shipment to CMS.

In addition to the modules listed above additional modules will be constructed at each assembly site using components without the full functionality. US CMS is responsible for these “dummy” components. US CMS is also responsible for purchase of all carbon fiber and aluminum carbon fiber parts for dummy, prototype, and functional modules to be built by US CMS. The responsibilities for assembly parts and fixtures are summarized in Tab. 4.5.

Prior to production, US CMS is responsible for:

- Development of assembly techniques and procedures (including jigs and fixtures) for the PS modules. Designs and documentation for PS modules will be provided to CERN for PS assembly in other countries.
- Validation of assembly techniques and procedures for 2S modules.

Additional US CMS responsibilities are as follows:

- The Al-CF material used for spacers and bridges is provided by a US company, and US CMS is responsible for the relationship with this critical vendor, as well as detailed studies to find the best way to machine Al-CF pieces (Sec. 4.5.2.4).
- US CMS is investigating different types of glues to evaluate the strength and thermal properties as well as gluing methods to provide a uniform thickness across sensor



modules.

- US CMS is investigating encapsulants with respect to viscosity, temperature required for curing, and radiation hardness for wire bonding encapsulation.
- US CMS is conducting studies of the use of materials (Parylene, Kapton) for PS modules and module designs to meet HV isolation requirements (Sec. 4.5.2.4.1).

Automation of module assembly is specific to the assembly site and will be done for PS and 2S modules by US CMS as well.

#### 4.5.2.1 Module construction

As described earlier, US CMS is responsible for construction of 3450 PS and 2300 2S modules.

There will be two module assembly centers within the US. Each center will receive module components and be responsible for inspecting, testing (in the case of hybrids and MaPSAs), and storing them prior to assembly. Each center will make its assembly jigs or procure them from designated providers that machine them to the standardized specifications. These jigs will be used for the sequence of assembly steps that build up the module from the components, which is largely achieved by gluing. When the mechanical assembly is complete, the

Table 4.5: Summary of responsibilities for module assembly parts and assembly fixtures.  
**FIXME:** how many MAPSA for bump-bonding? Check with Ron's section!

Assembly parts & fixtures	Responsibility	Comments
PS-s sensors	CERN	
MaPSA assemblies - MPA chips - PS-p sensors - Bump-bonding	CERN CERN and US CMS US CMS	Qualify vendors, provide MaPSA for all US CMS modules
2S sensors	CERN	
Al-CF spacers & bridges - raw material - machining	US CMS US CMS	Qualify vendor providing Al-CF for all CMS parts Provide parts for all US CMS modules
CF baseplates & stiffeners - raw material - machining	US CMS US CMS	Qualify vendor Provide parts for all US CMS modules
Flexible hybrids	CERN	
Dummy components	US CMS	Design dummy sensors & fabricate dummy sensors, blanks, dummy flex hybrids
PS assembly carriers	US CMS	Design and fabricate for US CMS modules
2S assembly carriers	US CMS	Fabricate for US CMS modules
PS assembly jigs	US CMS	Design and fabricate for US CMS modules
2S assembly jigs	US CMS	Fabricate for US CMS modules

center performs the wire bonding to connect the input channels to the read-out electronics, to connect to the pixel chips in the PS modules, and to make interconnections between the hybrids as needed. Each center will have test equipment for testing the hybrids as well as for testing the bonded modules. When a module passes the electrical testing, the wire bonds are then encapsulated for protection and the module is re-tested. More details about the needed equipment can be found in Sec. 4.5.3.

The construction of both PS and 2S modules is based primarily on manual jig-based assembly techniques. A number of other basic principles will also be used for the module assembly:

- The use of room temperature cure adhesives when possible, to avoid differential thermal coefficient based stresses.
- Standardized jig-based (the same design for all assembly centers) assembly with many identical jigs in each center for parallel production.
- Different jigs for each basic assembly step, allowing for simpler manipulations better suited for mass production.
- The most critical alignment, that of sensor to sensor, will be done by edge-based mechanical means which requires better than ten  $\mu\text{m}$  edge cut precision. The requirement is that individual strips in the top strip sensor be aligned to pixel strips in the bottom sensor to within 50  $\mu\text{m}$  (half the strip pitch).
- Uniform production site requirements, module assembly procedures and quality assurance planning.

The reason that a robot based module assembly, as was used in the existing CMS strip tracker, was not adopted is because the back-to-back sensor arrangement poses numerous problems for placement, precision alignment, and retaining the positioning during glue curing. The previous CMS tracker robot was based on a purely planar, single-sided sensor module which was intentionally designed for robotic assembly. The dual back-to-back sensors needed for the stub-based triggering for the upgraded tracker could not be easily adapted for robotic assembly. Even in the robotic assembly, a very large amount of human-based activities were needed: unpacking, inventory, visual inspection, component selection and placement, surveillance of the robot, inspection of assembly results, handling, testing, storing, and packing. It was determined that the additional manpower needed for doing the precision assembly steps previously done by the robot, given the right tooling and procedures, would not be excessive (less than 30% additional manpower). Robotic assistance will be used for some steps, however. For example, it is likely to be used in glue and encapsulant dispensing where it is very difficult to achieve a uniform and constant high quality result with manual dispensing. The use of robotics for other aspects of assembly is being pursued by US CMS.

#### 4.5.2.2 PS module construction

The assembly procedures for the PS modules are the same for the three module variants: 1.6 mm, 2.6 mm and 4.0 mm, which differ only by the thicknesses and configurations of Al-CF spacers. These are shown in Figure 4.34.

The PS module assembly consists of several discrete steps described below.

- Alignment and gluing of sensor pairs.  
Prior to assembly, all parts will be visually inspected for quality control. Al-CF parts will be evaluated using go/no-go gauges for correct thickness and flatness. Detailed

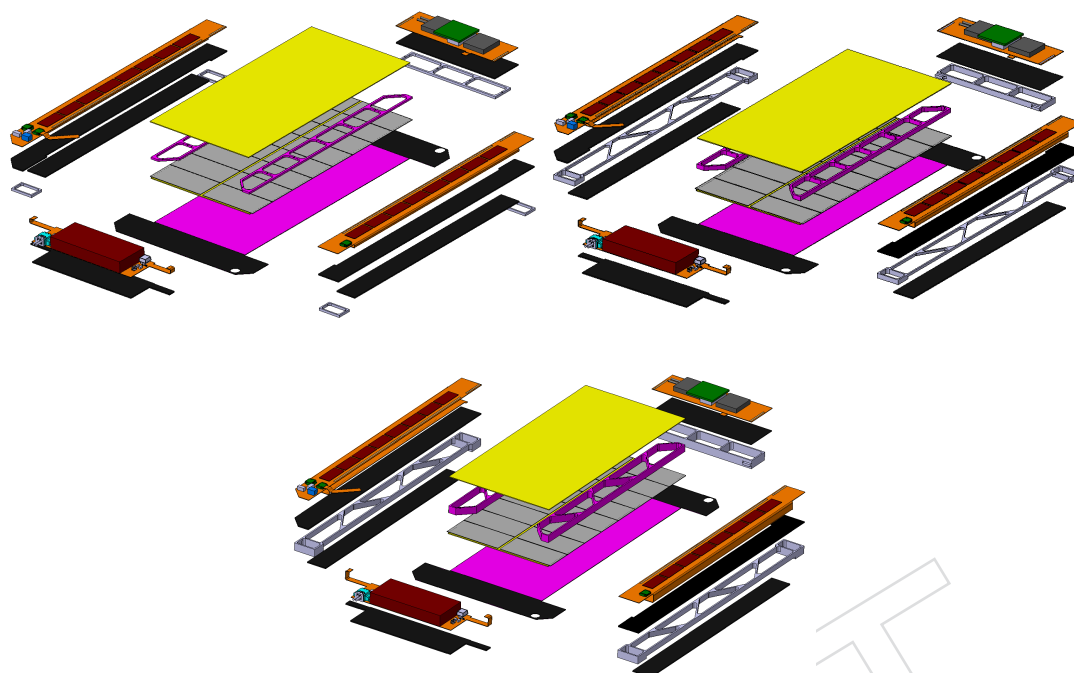


Figure 4.34: Exploded view of the 1.6 mm (top left), 2.6 mm (top right) and 4.0 mm (bottom) PS module variants **FIXME:** replace with fig with legend when available. .

QC for sensors will be done prior to receiving for assembly as described in Section 4.5.1. PS-s sensors will be visually inspected using automated image recognition software to evaluate the wire bond pads and strips. Wire bond pads on MaPSAs and flexible hybrids will be inspected using automated image recognition software as well.

The PS assembly starts with the gluing of the Kapton isolation foils and the HV bias circuit to the backplane of the PS-s sensors. A vacuum fixture will be used to pick up the very thin ( $25\ \mu\text{m}$ ) Kapton foils and hold them in position above the PS-s sensor. An assembly jig will be used to hold the PS-s sensor in position. A dispensing robot will be used to dispense a precise volume of glue on the back of the PS-s sensor. Once the Kapton foil has been placed on the sensor, the jig will be placed in a vacuum chamber to remove any air bubbles trapped between the foil and the sensor.

The next step is to glue the Al-CF sensor spacers to the PS-s sensor. A fixture picks up one or both Al-CF spacers and places them in a glue bath of a precise thickness. Robotic assistance will be used for this step to ensure a consistent thickness of glue across all PS modules. Once glue has been applied to the Al-CF spacer, it is positioned precisely on the backplane of the PS-s sensor with a dedicated assembly jig (Fig. 4.35). Methods are being investigated by US CMS to glue one Al-CF spacer at a time or both simultaneously.

The next step is to glue the MaPSA to the free side of the Al-CF spacers that have already been attached to the PS-s sensor. The MaPSA is placed in a dedicated assembly jig (Figure 4.35) with the MPA chips facing up. The PS-s sensor with spacers attached is pressed into a glue bath of precise thickness. Robotic assistance will be used for this step to ensure a consistent glue thickness. A fixture is used to lift the partial assembly out of the glue and position precisely on the MaPSA.

The PS-p sensor (part of the MaPSA) does not require isolation foils at this point since it will be glued to the CF base plate. However, it will require HV isolation since the backplane will face the CF. The base plate may be isolated with a parylene coating or may require a Kapton foil. In the case of a foil the foil may be glued to the CFRP base plate in a later step or to the sensor in this step.

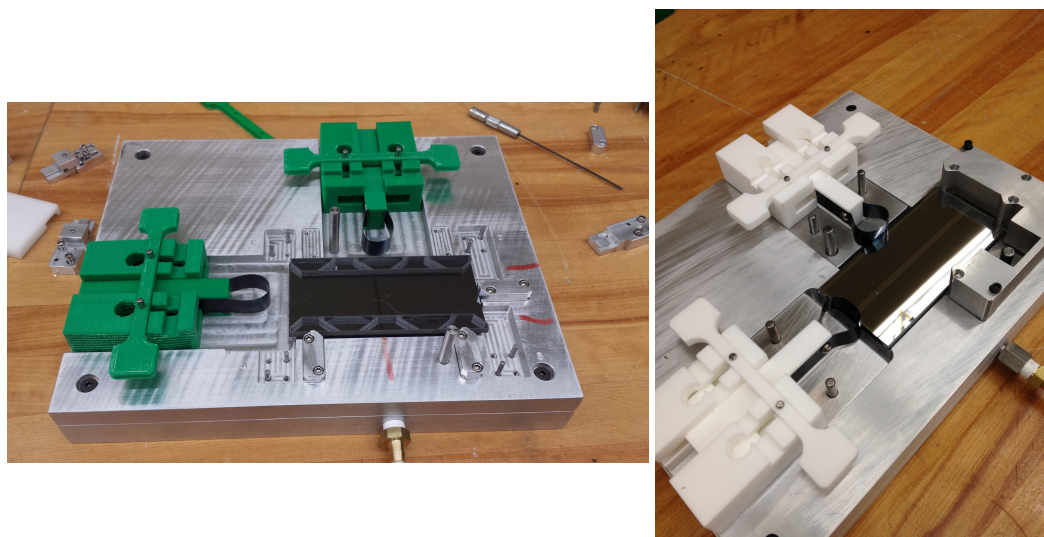


Figure 4.35: Images of the PS module jigs used to glue the PS-s sensor to the Al-CF spacers (left) and to glue the MaPSA assembly to the Al-CF spacers (right).

The basic structural support of the PS module is the CFRP baseplate since the sensor package and all the hybrids are glued to it. A module assembly jig (Fig. 4.35) holds the base plate in a precise position using the precision holes in the metal inserts of the plate. Guide pins and appropriate pick-up jigs align the sensor package and then the service hybrids and finally the FE hybrids all of which will be glued to the baseplate.

The challenging part for this assembly will be the glue dispensing, avoiding trapped air bubbles and achieving a thin but uniform coverage of the glue. It is likely that a low viscosity epoxy glue will be used for the sensor package and power service hybrid gluing to the base plate. This is because the surface area is large and achieving a thin, uniform coverage without bubbles will be very difficult with a high viscosity glue. In addition, the assembly jig with the module will be placed in a vacuum chamber to remove as much trapped air as possible. After this step is completed, the component assembly of the PS module is complete.

- Gluing of hybrids.

The flexible hybrid consists of a printed flexible substrate wrapped around an Al-CF hybrid spacer and a CF stiffener beneath the hybrid spacer. This will be received from the vendor already assembled. This is attached to the PS module with partial contact between the Al-CF hybrid spacer and the Al-CF sensor spacer along the long edge. The CF stiffener is exposed on either end and will be glued to the remaining portions of the CFRP baseplate extending beyond the width of the MaPSA assembly.

A dedicated gluing jig is used for gluing the hybrids to the PS module. Glue is applied to the zones of the Al-CF hybrid spacers that will be in contact with the sensor spacers. The FE hybrids are lowered in place and with a spring pusher, pushed against the sensor package so as to avoid a gap between sensor and hybrid which



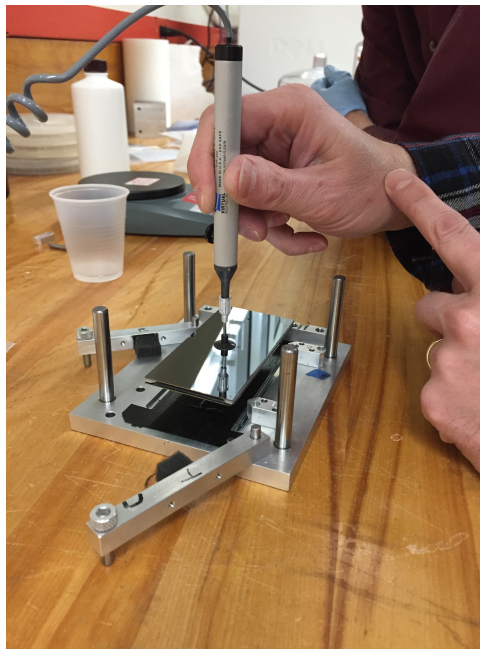


Figure 4.36: Photo of the PS module jig being used to glue the PS sensor assembly to the CFRP baseplate.

would make encapsulation of the bond wires difficult.

Finally, the opto and power hybrids which are roughly as wide as the PS sensors are glued directly to the CFRP baseplate next to the short edge of the MaPSA assembly. These are lowered in place using guide pins that match the holes in the hybrid. Weight bars are then placed on all hybrids so that the gluing zones of the hybrid have a uniform but thin layer of glue.

- Wire bonding and encapsulation

Once the glue is cured, the module is moved to the wire bonding jig. This jig uses 6 suction cups to safely hold the bottom PS-p sensor against the jig surface (covered with a clean room tissue). The jig has two hybrid support bars which can be raised and lowered using a screw system so that the bar will support the hybrid during bonding. Once placed on the jig, the vacuum can be activated which should pull down the module onto the jig surface and then it is ready for wire bonding the top side. Top side bonding will include the two rows of read-out connections between the PS-s sensor and hybrids, the top side ground bias connections, and the connections between FE and Service hybrids.

Once the top side bonding is finished, the module can be lifted off and turned over for the bottom side bonding. Note that the jig has been designed so that once bonded, the module can still be placed on the jig without damaging the existing wires so that repairs can be made.

After wire bonding, the module will go onto the PS module carrier plate which is designed to safely hold the module for testing, encapsulation and transport.

If the electrical testing of the module is successful, then the module can have its wire bond encapsulated. Wire encapsulation will be performed on a glue dispensing robot with a volumetric dispenser system. The goal is to completely encapsulate the row of bond wires with an elastic but tough transparent material with very little trapped air bubbles. The dispensing of the encapsulant is difficult owing to the fine



pitch of the wires which requires a low viscosity fluid and yet must not flow out to cover other areas of the sensor or hybrid which would imply a higher viscosity fluid. It may require two passes with different viscosity material to do this task. However, assuming an appropriate fluid can be found that allows for a single pass, then this step will first be performed on the top side of the module and after curing, performed on the bottom side.

After encapsulation, an electrical re-test will be made to check that no new problems have appeared. If the module passes this test, the assembly work is complete.

#### 4.5.2.3 2S module construction

US CMS assembly sites will use procedures and alignment fixtures (similar to those used for PS module assembly) developed by CERN and institutions outside the US documented below.

The assembly procedures are the same for the two variants of 2S modules, 1.8 mm and 4.0 mm, which differ only by the thicknesses and configurations of Al-CF spacers and bridges. These are shown in Figure 4.37.

The set of assembly jigs planned for 2S modules include:

- a jig for gluing of the Kapton isolation foils and HV bias circuit to the sensor backplane;
- a jig to glue the sensors to the Al-CF bridges, resulting in a sensor sandwich;
- a jig to glue the FE hybrids and service hybrid to the sensor sandwich;
- the wire bonding jig;
- a jig for encapsulation of the wire bonds.

In addition a module carrier is required as a handling device for electrical tests and for encapsulation. The first step of gluing Kapton foils and the HV bias flex circuits to the backplane of the sensors could be done in a separate location from the rest of the assembly and prepared in advance as soon as the sensors are available.

The most critical assembly step for alignment precision is the sensor sandwich gluing. The jig for this step is shown in Fig. 4.38 (left). This jig uses three precision machined stops to align the edges of the sensors, which must be cut to a high precision ( $< 10 \mu\text{m}$  error along the critical edge). After this step, a measurement of the alignment accuracy will be made [74]. Then the

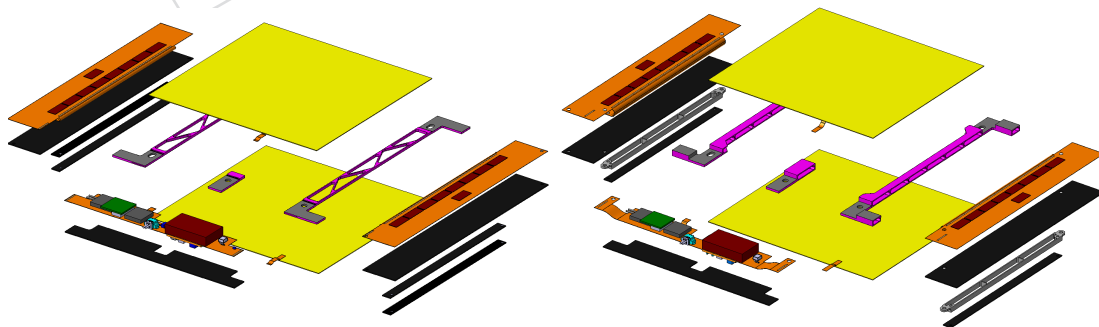


Figure 4.37: Exploded view of the 1.8 mm (left) and 4.0 mm (right) 2S module variants. **FIXME:** replace with fig with legend when available. Check TDR!

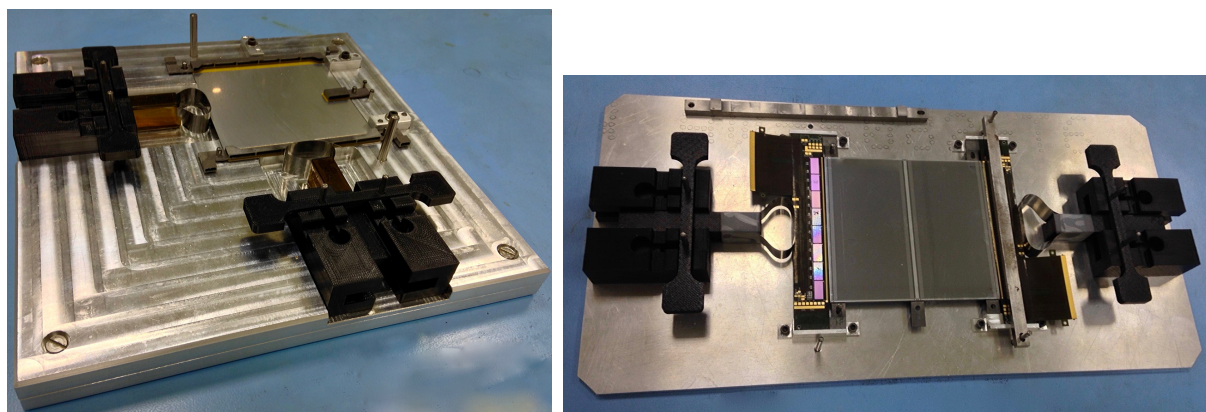


Figure 4.38: Images of the 2S module jigs used to glue the sensors to the Al-CF bridges (left) and to glue the front-end and service hybrids to the sensor package (right).

hybrids are glued to the sensor sandwich in another dedicated jig (Fig. 4.38 (right)). The gluing of hybrids is followed by the wire bonding, which will be performed first on the top side and then on the bottom side.

After wire bonding a full electrical test of the module will be performed to ensure that all strips are properly connected to the readout chips and that the readout chips and the service hybrid are fully functional. For this test the module will be moved to the module carrier, which is designed to provide protection to the module during the handling needed during testing. If the module passes the test, it will stay on the module carrier for the bond wire encapsulation and the final electrical acceptance test.

The module will remain on its own module carrier till it is installed on the mechanical structure.

Prototypes of “bare” 2S modules, consisting just of sensors and bridges, have been produced with both dummy and real sensors using the above assembly jig designs and procedures. After some tuning of the assembly jigs and the procedures, more than 12 bare modules were produced and all were within the required precision for the top to bottom sensor alignment (in particular, parallelism of the strips in the top and bottom sensor to better than 0.4 mrad is required). Figure 4.39 shows a dedicated setup used for the double-sided metrology of bare modules, along with an example measurement result obtained on that setup.

#### 4.5.2.4 Research & Development

Research and development efforts are ongoing specific to materials that affect assembly, performance, or reliability of PS modules. These include thermal testing as well as studies of machining reliability and high-voltage insulation of PS sensor spacers and baseplates. These studies are briefly described below.

- **Thermal performance:** PS module prototypes were constructed by US CMS to measure the thermal performance and to compare with results of similar tests done by CERN. A PS assembly was made using a 200  $\mu\text{m}$  piece of raw silicon, 2 Al-CF spacers, 3 layers of carbon fiber for a total of 200  $\mu\text{m}$ , and about 880 wire bonds. Heaters were used to simulate the heat load of a functional PS module. A custom built support structure with carbon foam, Rohacell foam, and a carbon fiber sheet was used with 3.2 mm cooling rods for thermal tests. The PS module was mounted in the support structure with re-workable Laird film. The cross-section of the PS module

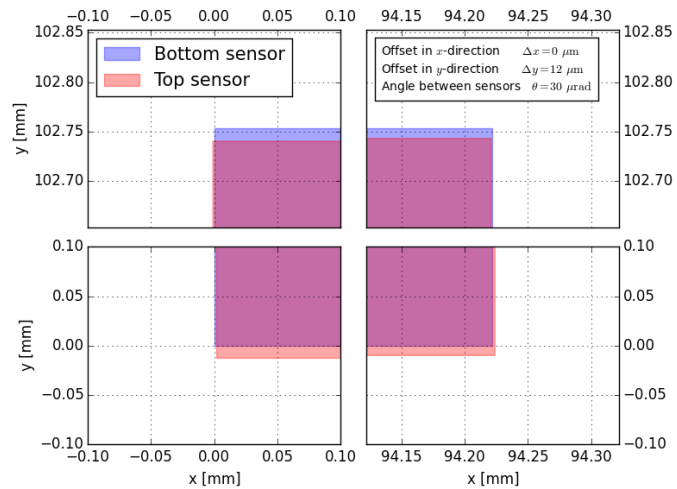
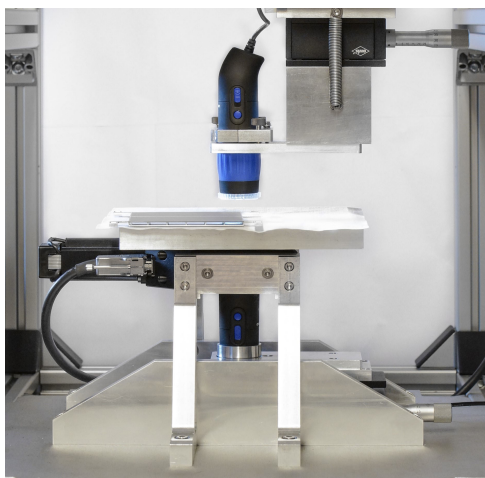


Figure 4.39: Left: photo of a prototype setup used to perform the double-sided metrology of bare 2S modules. The bare module is placed on a rotation table. Alignment marks on the module corners of the top and bottom sensor are photographed by the top and bottom camera, respectively. The measurements are referenced to each other by exploiting the fact that the axis of rotation is common. Right: results of a measurement. The four quadrants show zooms onto the corners. The coordinate perpendicular to the strips is denoted as  $x$ , while  $y$  is the coordinate parallel to the strips. The bottom sensor is shown in blue and the top sensor in pink. The result is reported in the legend; the precision is well within the specifications of  $\Delta x < 50 \mu\text{m}$ ,  $\Delta y < 100 \mu\text{m}$ , rotation angle between strips  $< 400 \mu\text{rad}$ .

prototype is shown in figure 4.40. Thermal measurements demonstrated a  $10^\circ\text{C}$  difference between the top and bottom RTDs (temperature detectors) which meets the design specifications and a linear response (i.e. twice the heat load gives twice the temperature difference).

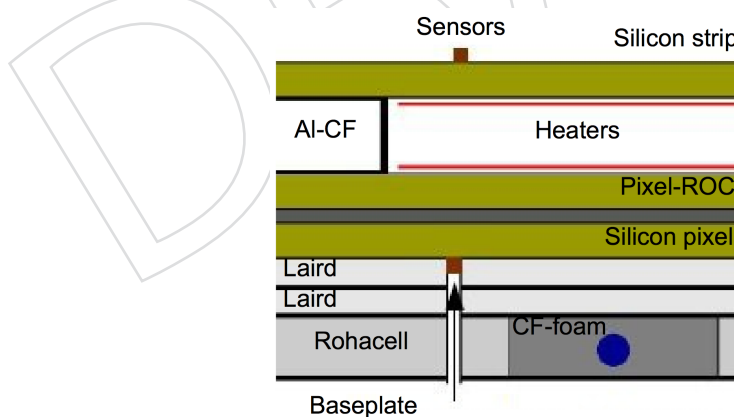


Figure 4.40: Cross-section of PS module prototype used for thermal testing.

- Al-CF spacers: The original baseline designs for PS and 2S Modules used Al-CF spacers with a Parylene coating to provide high voltage isolation. It was observed that a  $20 \mu\text{m}$  coating of Type C Parylene did not provide sufficient protection for 1 kV. In addition, for 2S Modules it was found that due to the fold-over of the FE hybrid on the sensor side of the hybrid, a thin piece of Kapton foil is needed between the sensors and the FE hybrid fold-over to prevent HV breakdown. For this reason,



Kapton foil has been incorporated into the 2S Module design between the Al-CF bridges and the sensors. For PS assemblies, on the other hand, since the fold-over is on the remote side of the FE hybrid, the use of a Parylene-coated Al-CF sensor spacer is still an option if it can be demonstrated to provide adequate protection.

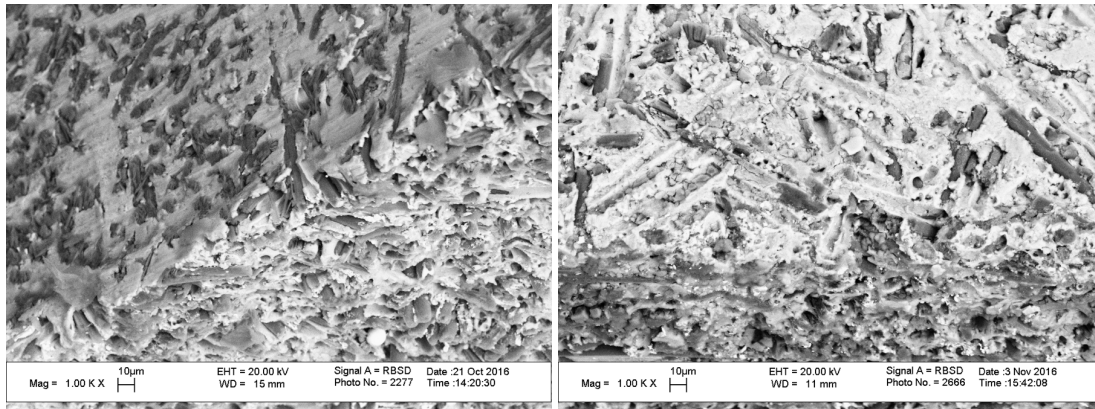


Figure 4.41: Electron microscope images of Al-CF spacer before (left) and after (right) O<sub>2</sub>-plasma etching. Images show significant reduction of carbon fibers which appear as darker images in surface of material.

- Plasma etching: The original tests with Parylene were assumed to have failed due to small carbon fibers protruding from the surface of the Al-CF spacers. Discussions with the Al-CF material vendor motivated a test using a oxygen-plasma etching process to oxidize the carbon fibers. A study was done using 100 W for 15 minutes on each side, and electron microscope images show a significant reduction in carbon fibers embedded in the surface (see Fig. 4.41).

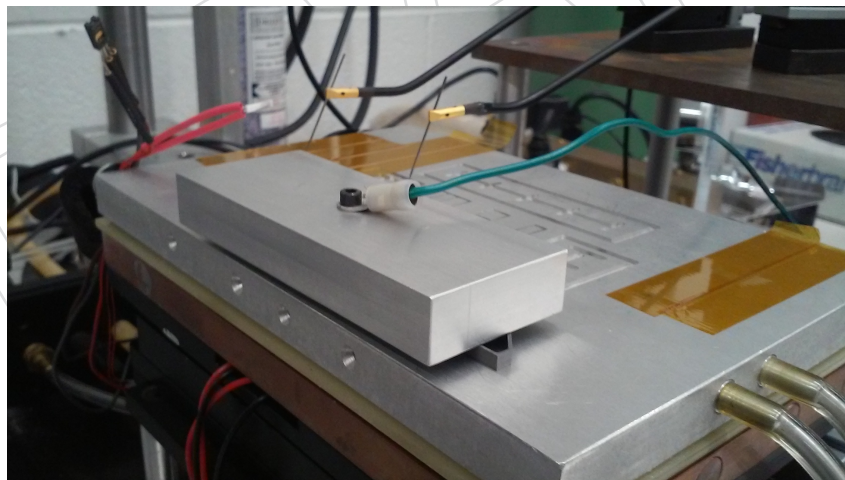


Figure 4.42: Test setup used to measure leakage current vs. high voltage for Al-CF spacer held between two aluminum blocks. One block was held at ground and measured the leakage current, while voltage was ramped up on the other block.

Studies were done with and without plasma etching for Al-CF spacers coated with Type N Parylene. Parylene-N has a dielectric strength similar to that of Parylene-C but a higher penetrating power. For this test, a 25  $\mu\text{m}$  coating of Type N Parylene was used. HV tests were done using a simple test setup described in Fig. 4.42. As shown in Fig. 4.43, this was demonstrated to hold up to 1 kV for spacers with and

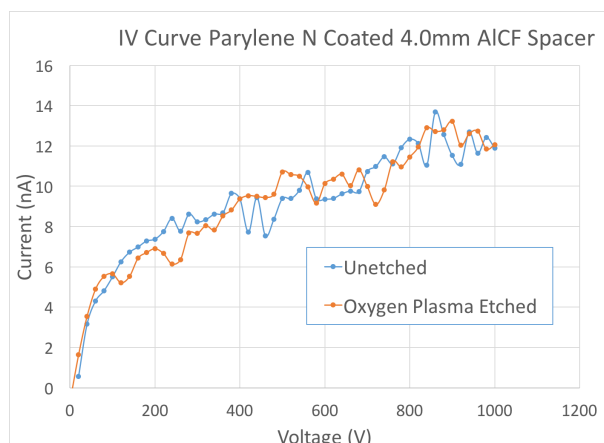


Figure 4.43: Leakage current vs. high voltage for Al-CF spacer using test setup described in Fig. 4.42. Two spacers coated with  $25\ \mu\text{m}$  Parylene Type N, one with and the other without plasma etching prior to Parylene coated, held voltage up to 1000 V.

without plasma etching. Studies are ongoing to demonstrate the feasibility of large-scale use of Al-CF spacers coated with Parylene-N for PS Modules.

- Annealing: Al-CF spacers for the 1.6 mm PS Modules are very thin (1 mm) and were found to bow as a result of the normal machining process. The Al-CF material vendor suggested the use of annealing to remove the bowing after machining. Studies demonstrated for several Al-CF pieces that annealing at 300 C for 30 minutes with Al-CF spacer held between graphite blocks improved the flatness from  $500\ \mu\text{m}$  (as measured with shim inserted under the bowed spacer on flat surface) to  $< 40\ \mu\text{m}$ .

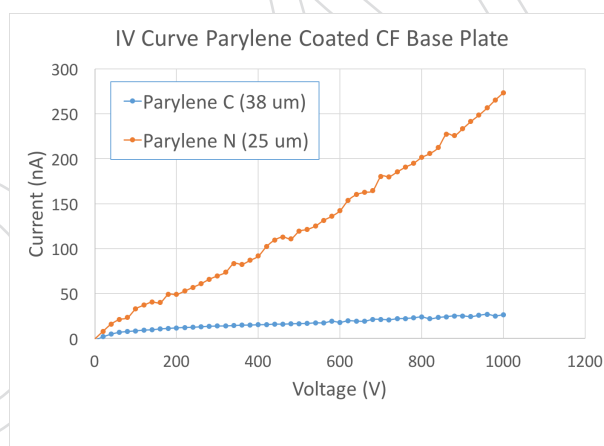


Figure 4.44: Leakage current vs. high voltage for two CFRP baseplates held between two aluminum blocks. One aluminum block was held at ground and measured the leakage current, while voltage was ramped up on the other aluminum block. One of the CF baseplates was coated with  $25\ \mu\text{m}$  Parylene Type N, and the other was coated with  $25\ \mu\text{m}$  Parylene Type C. For the 1st two tests (labeled “FullCenter”), the aluminum block was larger than the baseplate. For the remaining tests, a small  $1'' \times 1''$  block was used to measure the leakage current. In all cases voltage held up to 1000 V without breakdown.

- Parylene coating: As for Al-CF spacers, in the original design of PS modules the CFRP baseplates were coated with Parylene to provide protection against high voltage. Studies have been done demonstrating the effectiveness of a  $25\ \mu\text{m}$  thickness of Parylene Type N or Type C up to 1000 V (see Fig. 4.44). More studies are needed to



demonstrate the feasibility of using Parylene on a large scale.

- Co-curing CF with Kapton: US CMS is investigating the possibility of using carbon fiber sheets co-cured with Kapton for production of CFRP baseplates and stiffeners where high voltage isolation is needed.

#### 4.5.2.5 Functional tests during the production phase

Test systems along with QA/QC protocols are developed centrally by the collaboration and adopted by the various module assembly centers and vendors, where applicable. The following tests are foreseen for the production phase:

- MaPSA: described in Sec. 4.5.3
- Front-end hybrids: Basic testing will be performed at the vendor in order to maximize the yield of received hybrids. These tests are expected to include ten passive thermal cycles followed by a simple functional test (read/write of the ASICs' configuration registers). Upon reception by the assembly centre the assembled hybrids will be weighed. The weight control allows to quickly detect process changes that may happen during manufacturing or assembling steps, which can potentially affect the reliability of the final product. Functional tests will be performed at room temperature and at the operation temperature. For the qualification of the 2S FE hybrids those tests include the following qualification steps:
  - determination of the power consumption,
  - verification of the ASICs' configurability,
  - full calibration of the front-end ASICs,
  - testing of the analogue input channels, i.e. test of the connectivity of the wire bond pads to the ASICs, verification of the connectivity of the fast data lines, diagnosis of input channel shorts.

The specified list of parameters to be tested, the implemented testing methods, the way hybrids are temporarily connected to the test setup and the readout chain will be similar in the production test system.

- Service hybrids: All service hybrids will undergo testing at the vendor. This is expected to include automated optical inspection, X-ray imaging of the chip areas (at least on a sample basis), passive thermal cycling, and functional tests. Tests on a sample basis will be performed at the institute responsible for the procurement. All service hybrids will be subjected to a functional test, both at room temperature and operating temperature, at the module assembly centres before being assembled into modules. The following tests are foreseen:
  - verification of the dimensions,
  - check of the functionality of all active devices (DC-DC converters under maximum load, LpGBT, VTRx+),
  - verification of the low voltage and bias voltage distribution,
  - validation of the data path between FE hybrid and LpGBT and between LpGBT and VTRx+,
  - validation of the I<sup>2</sup>C bus,
  - check of connectivity for the clock, reset, and trigger lines,

- readout of the slow control data, including the temperature sensor.

- Single module tests: Two connectivity and functional tests will be carried out. The first test will be done after wire bonding; the second after wire bond encapsulation. These tests provide an immediate feedback on missing, damaged or shorted wire bonds. The module will be tested using a DAQ system based on the FC7  $\mu$ TCA board. The test will be performed at room temperature.
- Multi-module tests: Sets of about ten modules will be operated together for a few days with temperatures cycling between room temperature and operating temperature. A DAQ system based on the FC7  $\mu$ TCA board will be used for this test. This test should identify early failures of the assembled modules both in terms of mechanical and electrical problems. The US CMS is responsible for developing the burn-in system for CMS. Initial drawings for the enclosure can be seen in Fig. 4.45. Modules mounted on their carriers will be inserted into designated slots and clamped to the the support to ensure efficient heat conduction between the module and the cold plate. The modules are readout by engaging dedicated connectors on carrier plates into a connector built into the enclosure's backplane.
- Sub-assembly tests: Modules mounted on large structures will be operated for a few days with temperature cycling between room temperature and operating temperature. These tests will validate the modules' functionality, the quality of the module cooling contacts, and the uniformity of the temperature across the structure. The sub-assembly will be tested at the operating temperature, with cooling provided by a dedicated CO<sub>2</sub> system.

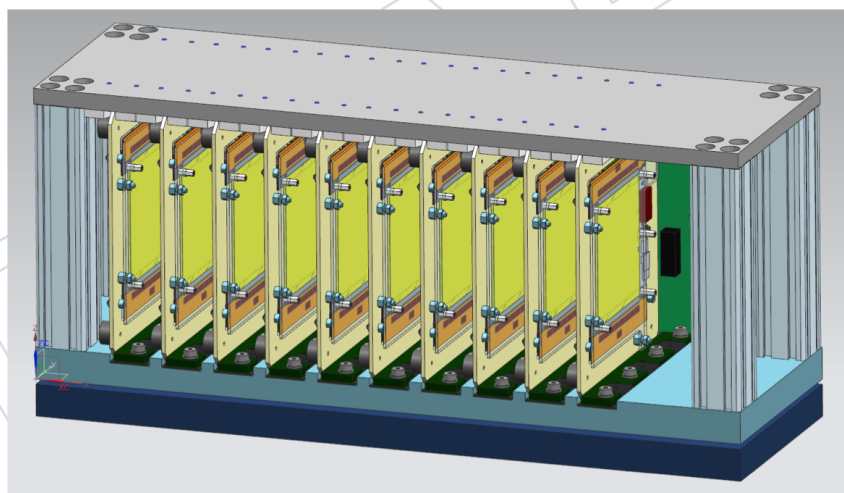


Figure 4.45: Preliminary burn-in system based on Peltier elements.

### 4.5.3 Electronics

US CMS responsibilities include:

- Vendor selection for MaPSA bump-bonding and QA/QC of all bump-bonded MaPSAs
- Contribution to the development of the DTC back-end board.
- Development of the Off-the-shelf-data-acquisition-system (OTSDAQ).

Details about the MaPSA and the DTC board can be found in Sec. 4.4.3, while the OTSDAQ is presented below.

#### 4.5.3.1 MaPSA assembly, process development, testing

US CMS will deliver the MaPSAs for the entire OT **FIXME: we must check this statement!.** The OT has 5332 PS modules. With 15% spares and 5% pre-production prototypes this totals 6480 MAPSAs. There are 16 MPA chips per MAPSA totaling 102,400 chips. US CMS is responsible for the bump-bonding vendor selection and qualification, and for the development of the QA/QC test system and test protocols. The latter be used at the vendor and at the various assembly centers in the US and abroad. These activities are carried out by Fermilab, Rutgers University, and Brown University.

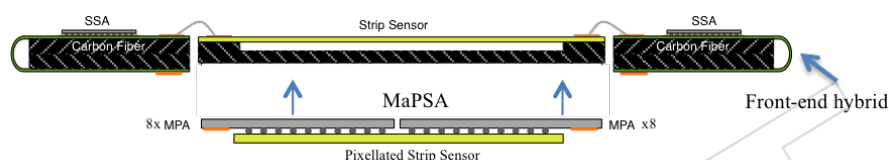


Figure 4.46: Edge view of the PS module showing the position of the MAPSA module in the assembly.

The R&D and pre-production work includes developing full size electrically simplified (dummy) assemblies with geometry and inter-silicon gaps similar to those of the final MaPSAs. These dummies will be used to qualify the dielectric strength of the underfill chosen for the assemblies. The outer edges of the n-on-p sensors anticipated for the MaPSA assemblies are at the applied sensor bias voltage (up to 800V). Because the MPA chips that overhang these edges (Fig. 4.46) are at ground potential there is significant danger of breakdown across the  $\approx 80\mu\text{m}$  gap between sensor and readout chip. For this reason the development of the MaPSA assembly process includes the fabrication of 60 dummy structures. The dummy MaPSA assemblies will have daisy chain patterns to verify bump connectivity. Edge underfill will be applied and tested to confirm that the high voltage breakdown exceeds 800 V.

An initial set of dummy MaPSAs has been fabricated and undergone probe testing. Figure 4.47 shows an image of one of the initial modules undergoing probe testing. These tests demonstrate our ability to identify problems in the bumps or scratches in the starting material. Subsequent versions will incorporate serpentine resistors to allow us to verify the quality of the patterned silicon before bump bonding.

The pre-series parts (268 assemblies) will instead utilize final or near-final versions of the MPA chips and sensors. The pre-series assemblies will be constructed to fully verify the quality of the assembly, including sensor and chip yield. Final verification of the assembly process will include thermal cycling and environmental testing of both the dummy and pre-series assemblies.

During the production phase US CMS will receive MPA ASICs (with bumps on them) from CERN and PS-P sensors from US and Europe and will work with vendor(s) to produce the MaPSAs to be used for module construction in the US and abroad. These parts will have under-bump metallization and bumps applied, will be thinned to the final thickness and diced. The MaPSA vendor will bump bond the MPA chips to the sensors, apply under-fill and test the

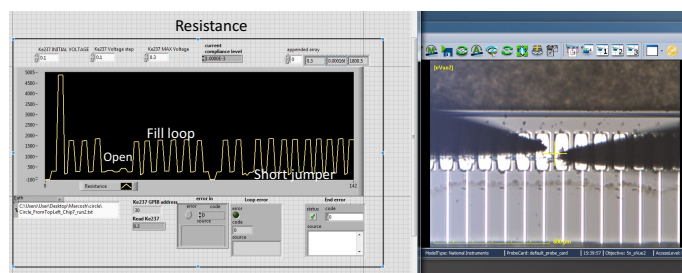


Figure 4.47: Image of a dummy MAPSA undergoing tests on an automatic probe station. The pad pairs alternate between short jumper connections and higher resistance loop connections through a daisy chain. An open region caused by a scratch has been identified

resulting assembly. These tested assemblies will then be sent to the module assembly centers.

Both the vendor and the assembly centers will test the MaPSAs using the system developed by US CMS. This is based on a probe card (developed by Fermilab in collaboration with Rutgers University) and an interface card (developed by Rutgers University in collaboration with Fermilab). All chips in the assembled MaPSA devices will be probe tested to verify operation of the 16 chips and the connectivity of the bumps. We expect to adapt hardware and software used for the on-wafer MPA tests for this purpose and transfer the test equipment to the MaPSA vendor. The testing will also include verification of sensor functionality after bump assembly and the dielectric strength of the applied underfill. The preliminary set of QA/QC tests includes:

- Measurement of the current as a function of the sensor bias voltage, to assure that the sensor was not damaged during bump bonding or transportation.
- Measurement of noise. A noise scan of the MaPSA provides a detailed map of the disconnected pixels. Disconnected pixels show a noise very close to the value before the assembly, while connected pixels show a noise increase of around 25%. This effect is enhanced when no sensor bias is applied.
- Measurement of the hit map when the device is exposed to X-rays or a radioactive source. This measurement provides the most reliable method to obtain a map of the disconnected pixels, however it can only be done on a sample basis. In this case the sensor must be biased during the test.

We expect to require:

- Bump connectivity yield to be greater than 98%.
- All 16 chips fully functional.
- The assembly must hold off the final full depletion voltage (800V).

#### 4.5.3.2 Data-trigger-control back-end board

The plan for the DTC back-end board in CMS is not yet formed. However, since the performance of OT as a whole critically relies on its performance, US CMS is committed to its development. The DTC will be developed by the consortium of US and non-US institutions. Currently the US institutions involved in the project are Fermilab, Rutgers University, and Boston University.

Boston University will develop an ATCA test module to test the performance of 25 Gbps back-plane links in an ATCA shelf. This module will contain a Kintex Ultrascale+ FPGA (likely a XCKU5P) at least two four-pair bidirectional lanes routed to the zone 2 data transport zone on the ATCA backplane. One lane will allow for testing of data transfer from the test board to a switch blade located in one of the ATCA dedicated switch slots, and another will allow for testing of blade-to-blade communication across the entire length of the ATCA backplane. The module will also contain a Xilinx Zynq processor for Ethernet communication and control.

Fermilab will develop common functional firmware blocks using the Pulsar IIb ATCA board [75] as hardware platform for developments of the OT DTC. The firmware development would focus on (1) integrating existing blocks (for example, from FC7 firmware) as well as (2) providing new blocks as needed. This development would support OT system test activities by augmenting existing firmware in the Pulsar IIb FPGA with (for example) detector configuration, fast controls, and clock distribution blocks. The system test development would build from the system test work of the Versatile Link Plus project carried out by Fermilab up until now. For example, firmware to provide Pulsar IIb support for specific protocols that might be implemented in the lpGBT10 ASIC would extend the boundaries of the system test environment. Rutgers University will work in close collaboration with Fermilab on the FW development.

#### 4.5.3.3 Off-the-shelf DAQ

US CMS is responsible for delivering the Off-the-shelf-data-acquisition-system (OTSDAQ [76]) to CMS. **FIXME:** need a paper for OTSDAQ!!

The OTSDAQ is a generic data acquisition system based on XDAQ and developed by the Fermilab Computing Division. The XDAQ itself is the official framework adopted by the CMS online DAQ group. US CMS will develop and integrate the software used for testing (middleware) into OTSDAQ and in turn will integrate OTSDAQ into the central DAQ system of CMS for online data taking. This activity is led by Fermilab with contributions from Rutgers University.

Given the scalability of this DAQ, US CMS will first adapt it to support test beams and test stands, on the different  $\mu$ TCA and ATCA boards, for modules production, making sure that it will scale smoothly to be adopted as the DAQ for the outer tracker detector in CMS. The development of OTSDAQ therefore includes the following sub-projects:

- OTSDAQ for testing prototypes at beam facilities and laboratories. The integration of the middleware into OTSDAQ has already started for both the 2S and the PS prototypes and a first release can be used already in beam tests.
- OTSDAQ for testing production components (e.g. MaPSAs, modules). At the time of writing, work is on-going to coherently integrate the calibrations needed to characterize the modules by re-configuring OTSDAQ to operate both on a single module and a multi-module system.
- OTSDQ for central data acquisition.

#### 4.5.4 Flat Barrel (Flat TBPS)

US CMS is committed to design, prototype, produce, assemble, and deliver the three TBPS layers where the modules are placed at zero-tilted angle (Flat Barrel TBPS). The reader is reminded



here that the Flat TBPS consists of three layers of modules arrayed on planks that provide cooling and support. It consists of a total of 952 PS modules mounted on 80 planks. Table 4.6 provides the basic geometry of the three layers.

Table 4.6: Geometry of the flat barrel section. R values are in mm.

Layer	Planks per Layer	Modules per Plank	Rmin	Rmax
1	18	7	205	216
2	26	11	333	344
3	36	15	485	497

While a general introduction about the inner flat barrel or Flat TBPS is provided in Sec. 4.4.4, additional details are provided in what follows.

Modules in the Flat TBPS section will be supported by planks which also provide cooling and alignment. Modules are attached to both the tops and bottoms of the planks. The overall thickness is minimized to allow for minimal spacing in Z between the top and bottom module sets. The plank consists of top and bottom carbon fiber facings spaced by high thermal conductivity carbon foam. The foam provides both rigidity and acts as an efficient heat spreader. Stainless steel cooling pipes are imbedded in trenches milled into the carbon foam and lined with thermally conductive epoxy to provide a smooth surface for heat transfer from the cooling pipe to the foam. The tubing is then imbedded in the foam with thermal grease and the top facing is glued to the sandwich. The total thickness of the assembly is 3.5 mm. Precision titanium inserts are embedded to provide support and alignment points for the modules. Modules are held on the plank by a set of screws as well as a layer of phase change adhesive similar to the adhesive being used for the CMS forward pixel detector. Inserts in the plank provide alignment and initial positioning, the phase change adhesive provides a large area thermal path between the module and the plank support structure. The average radiation length of the plank structure is 0.42%. Figure 4.48 shows a drawing of the inner layer plank with 4+3 modules mounted on the two sides.

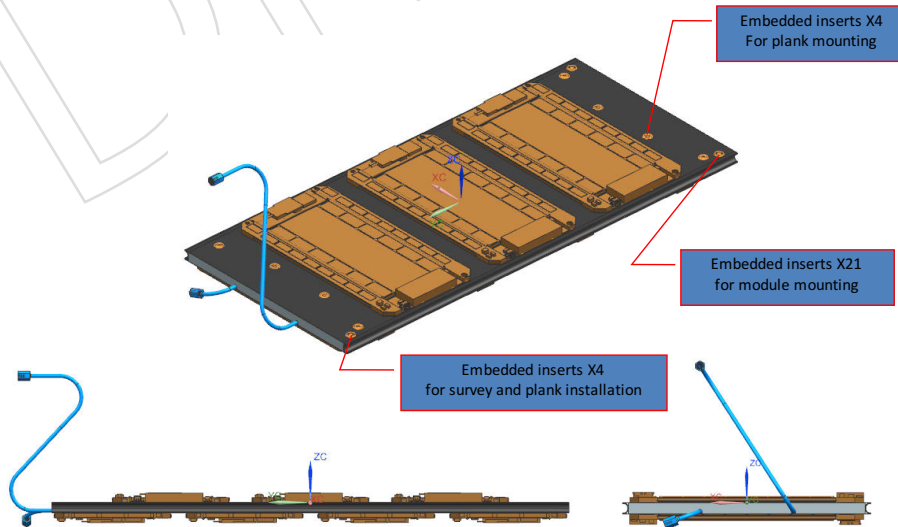


Figure 4.48: Drawing of modules mounted on the plank for layer 1 of the tracker inner flat barrel. Cooling tubes are bent to mate with neighbor planks in phi.

Planks on each of the flat barrel layers are supported by a pair of end rings. These rings provide both support and precise positioning of the two plank sublayers. Each ring consists of front and back carbon fiber facings separated by carbon or airex foam. The two rings are spaced by a set of support beams at the outer radius that also provide support for cooling lines and signal cable routing. A pair of rings is required for each of the three flat barrel layers. The inner sublayer planks are secured to precision inserts glued to a carbon fiber inner ring. Outer planks are secured to the support rings by precise block inserts glued to the inner and outer facings. Figure 4.49 shows the design for the inner support ring in it's assembly fixture.

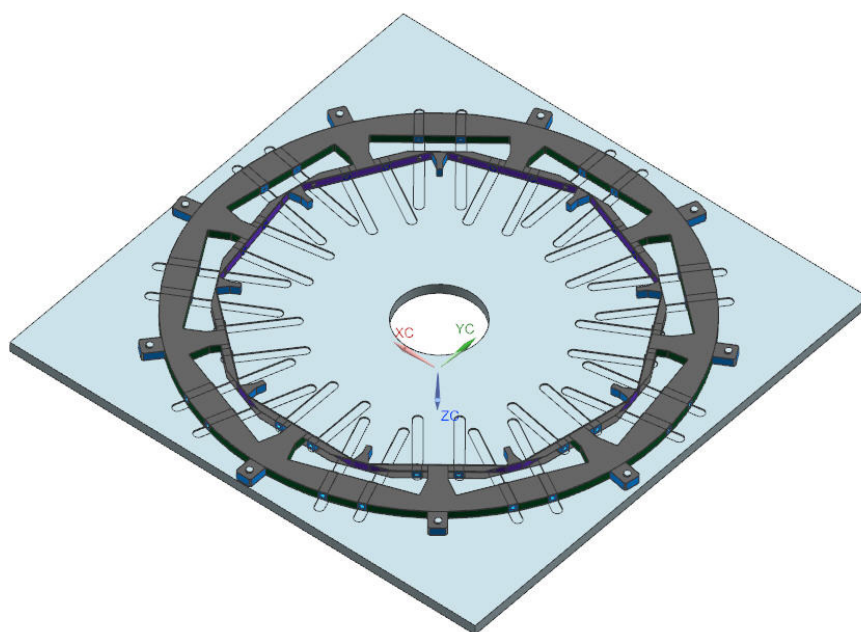


Figure 4.49: Carbon fiber/foam support ring for the inner layer of the flat barrel on it's assembly tooling plate. Locations of the support inserts can be seen in blue. Fingers projecting to the inner radius are "bumpers" to prevent damage during insertion over the pixel support tube.

All planks are cooled by dual phase  $\text{CO}_2$  circulated throughout the outer tracker structure. One cooling loop serves 2 planks connected in series with supply and return on the same side. In the current design the pipes are 2.2 mm outer diameter stainless steel with 0.127 mm walls. We will use connector assemblies similar to those currently being utilized for the CMS FPIX detector. The tubes are preformed to connect to adjacent modules and strain relieved at the support beams. Forces imposed on the pipes are limited to insure that there is no excessive force on the plank. This will be checked during the initial assembly.

Services include power cabling to the DC-DC converters and optical fibers from the module GBTs. These will be routed along the edge of the plank and then up to the support beam. They then are routed along the support beam to the end of the TBPS assembly where they are routed vertically out toward the end of the tracker.

#### 4.5.4.1 Assembly

The flat barrel section can be fabricated and tested independently. After testing it will be coupled to and supported by the two outboard tilted barrel sections. Services and cables are routed from the inner flat barrel along the support beams to the ends of the barrel section. Beams in

the flat barrel region are coupled to similar beams in the tilted sections to form a contiguous support structure. The entire PS inner barrel is supported from its ends by the array of support beams.

Once the flat barrel section is completed and tested it will be inserted by sliding the assembly over the inner support tube. This tube will also support the pixel barrel detectors from its inner surface. The tube is assumed to be fabricated as a double walled carbon fiber cylinder with hexagonal core inner spacers.

Testing of partial barrels with full readout will proceed at Fermilab. The barrels will then be disassembled and the parts shipped to CERN.

We expect final assembly of the flat barrel layers will be done at CERN and final tests of fully populated barrels will be performed at CERN as well.

Assembly jigs and fixtures will be provided by the US. Partial barrels will be assembled and tested at Fermilab to develop and test these fixtures and for mechanical quality control as well as cooling tests and electrical system testing. The conceptual design of the flat barrel support is quite similar (two spaced rings) to that used for the D0 silicon tracker. In that case the two rings were aligned and then supported by a precision rotation stage during assembly. An adjustable sliding stage was then used to insert ladders. A similar concept can be applied on a larger scale to the positioning of planks on each layer of the flat barrel section. We are in the initial stage of this design.

#### 4.5.4.2 Quality control

Critical positioning of modules and planks will be set by precision features in the support structures. These in turn are set by precision machining of jigs and fixtures or by assembly under a Coordinate Measuring Machine (CMM). The quality of these parts will be validated using the CMMs at SiDet at Fermilab. Modules are placed onto planks utilizing the positioning inserts embedded in the plank structures. Thermal contact is made using a phase change adhesive (Laird PCM-583). This adhesive was used for the CMS FPIX detector, has high thermal conductivity and is reworkable. Each assembled plank will be measured before and after module installation under an optical CMM to record sensor positions and validate the quality of the full mechanical assembly. Each plank will be tested and burned in with full readout at nominal operating temperature before being qualified for assembly into a barrel.

#### 4.5.4.3 Research and development

The plank concept has undergone both finite element simulation and prototype testing in the Fermilab dual phase CO<sub>2</sub> test facility. Figure 4.50 shows a cross section of a thermal model of a module mounted onto a carbon fiber/carbon foam plank. The critical temperatures are of the two sensors. Finite element thermal models were used to understand heat flow and thermal paths. Prototype tests (Figure 4.51) were then used to refine and confirm the thermal models. This led to the conclusion that the junction between the small area tubing carrying the CO<sub>2</sub> coolant and the carbon foam is a critical thermal interface. The carbon foam, although very thermally conductive, presents problems for making contact to other materials due to the fibrous surface. Our current technique consists of lining the trough in the foam with boron nitride loaded epoxy, machining that surface, and then imbedding the tubing in thermal grease.

A possible alternative would be to glue the tubing directly to the foam trough, letting the foam fibers flex to take the difference in thermal expansion coefficient between the carbon foam and the stainless steel.

We have also tested prototype planks of different thicknesses to measure the gravitational deflection. These planks were loaded with the expected weight of the PS modules and measured for various lengths and types of support. A summary of these tests are shown in Tab. 4.7. The results of the tests are consistent with both finite element and hand calculations and give us confidence that we have a good understanding of the performance of the plank structures. We plan to use planks of 3.54 mm thickness for layers 1 and 2 (0.42% radiation length) and 4.35 mm thickness (0.5% radiation length) for layer 3.

Our tests indicate that the plank design meets both thermal and mechanical performance requirements.

Table 4.7: Results of prototype mechanical tests

Thickness (mm)	Support dist. (mm)	Calc. Defl. ( $\mu\text{m}$ )	Meas. Defl.
3.54	256	9	10
3.54	415	30	49
3.54	623	199	232
4.35	256	6	6
4.35	415	23	31
4.35	623	148	170

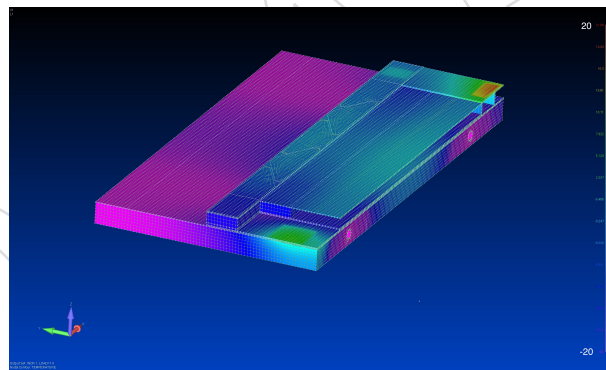


Figure 4.50: Thermal model of a module mounted on a plank. The coolant is assumed to be at -20 degrees C. Hot spots on the edges correspond to the GBT and DC-DC converter assemblies



Figure 4.51: Prototype module mounted on a test plank inside a cold box at SiDet. Multiple heaters are mounted on the module to simulate heat sources. An array of RTD temperature sensors are mounted throughout the assembly.

Table 4.8: Summary of MaPSA-Light assemblies production and quality. (*A1=Aemtec, A2=Novapack, A3=Leti, A4=CVI*)

Assembler	First Campaign		Second Campaign	
	made/tested	comment	made/tested	comment
A1	5/4	bad bump connectivity	15/8	good
A2	5/4	good	7/1	one out of 6 chips dead
A3	6/5	chip damage		
A4			12/1	metallurgy problems

### 4.5.5 Prototyping, system tests, and beam tests

#### 4.5.5.1 Prototypes

The US CMS has taken charge of the testing of the first ASIC prototype for PS module, MPA-Light. The MPA-Light has the same FE analog circuitry as the MPA ( $16 \times 3$  channels), and can be bump-bonded to a sensor similarly to full MPA. This provides a way to extensively test the ASIC response to ionization in silicon sensors. Small versions of PS-p sensors with  $48 \times 6$  pixels (to be bonded to six MPA-Light ASICs) were procured on 4 inch wafers. Diced MPA-Light chips and sensor wafers were submitted for bump-bonding to four vendors in two production campaigns. The test results from the first campaign were used to refine the bump-bonding process for the second one. In particular, assembler A1 had drastically improved the assembly quality (see Tab.4.8).

ASIC and sensor prototypes become available at different times, and need to be tested and verified, before the entire electronics chain is available. That necessitates design and deployment of multiple test systems, both for validation of digital circuits and for testing module prototypes with particle beams, lasers, and radioactive sources. The above mentioned tests were carried out by connecting the MaPSA-Light to a carrier board and an interface board entirely designed by US CMS. These systems are the predecessors to the final setup for testing production MaPSAs US CMS is committed to deliver (mentioned in Sec. 4.5.3).

#### 4.5.5.2 System tests

US CMS tested a number of prototypes: MaPSA-Light assemblies, micro-modules, 2S mini modules, and 2S full size modules. The test systems for a single MaPSA-Light and for a the micro-module were developed by US CMS while that for 2S mini module and for the 2S full size module by CERN. Fig. 4.52 shows that the time walk of the MaPSA-Light measured using the laser station at Sidet (Fermilab) is consistent with the specifications. US CMS also carried out functional tests of one 2S mini module built by CERN (described in Sec. 4.4.5.1). Figure 4.53 shows the noise as a function of time when the device is operated at -30 C and the chip configuration is adapted to operate at this low temperature. The 2S mini module is stable over a period of time corresponding to approximately one week. The three full size 2S modules (described in Sec. 4.4.5.1) built by CERN were tested at room temperature. The noise is presented in Fig.4.54. At the time of writing a detailed investigation of the modules' performance is on-going. These preliminary results nevertheless indicate that the noise is compatible with the specifications<sup>5</sup>.

<sup>5</sup> Preliminary measurements of the absolute calibration indicate that one  $V_{cth}$  corresponds to approximately 300-400 electrons



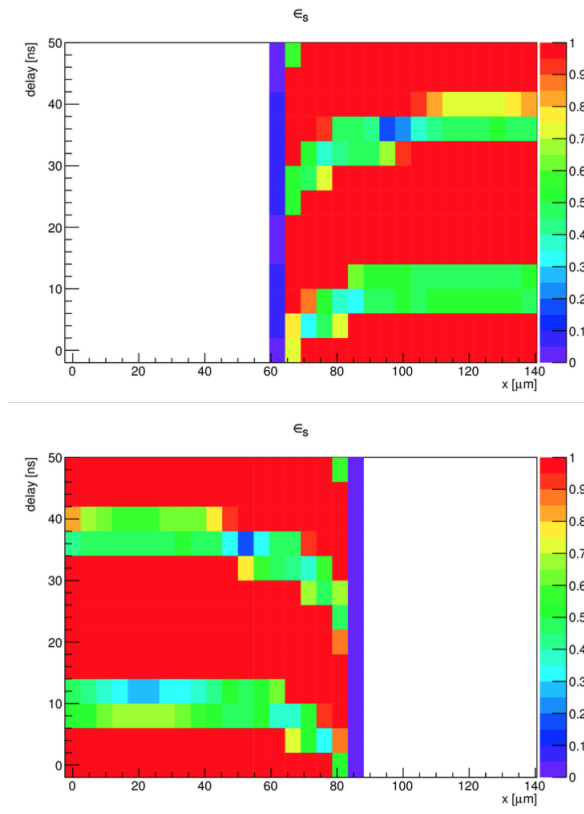


Figure 4.52: Laser hit efficiency in two neighboring pixels as a function of phase between the laser pulse and MPA-Light clock (Y-axis) and the coordinate of the laser.

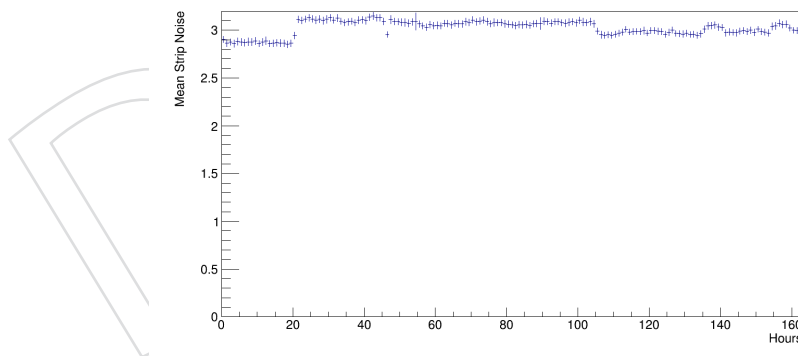


Figure 4.53: Noise of the 2S mini module operated at -30 C. The mean and RMS of the distribution are shown. **FIXME:** Plot will be updated.

3343

### 3344 4.5.5.3 Test beams

3345 The MaPSA-Light assembly was characterized using the Fermilab beam facility providing  
 3346 120 GeV protons. The response of the MaPSA-Light assembly was determined by measuring  
 3347 the hit reconstruction efficiency as a function of the phase difference of the MPA clock and the  
 3348 beam clock. The result shown in Fig. 4.55 indicates that the hit efficiency is close to 100% and

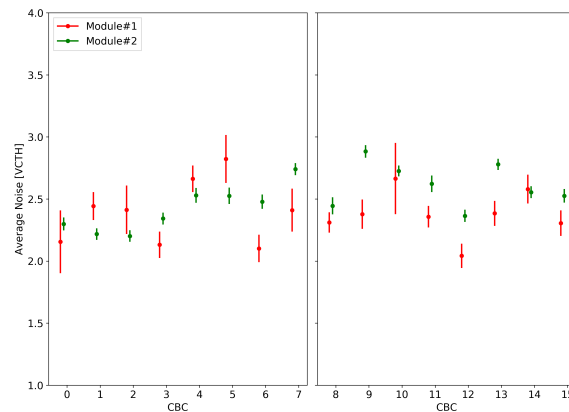


Figure 4.54: Noise of the three 2S full size module operated at room temperature. Average noise for each of the 16 CBC's on all 3 modules. The uncertainties are given by the RMS of the measurement. The noise is derived by fitting a cumulative distribution function of the normal distribution to the S-curve and the width is taken as the noise. **FIXME:** Plot will be updated. Module 3 is missing

3349 the time walk is consistent with the specifications.

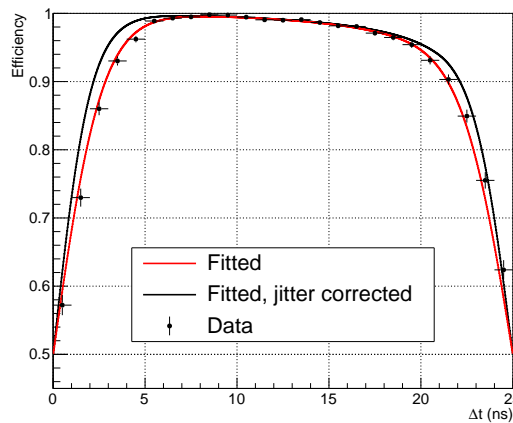


Figure 4.55: Hit efficiency as a function of the phase difference between the particle arrival time and the 40 MHz clock edge in the MPA. The measured efficiency spectrum is shown by the black dots, the red curve represents a fit to the measurement, and the black curve shows the jitter corrected efficiency distribution.

DRAFT

## Chapter 5

# Barrel Hadronic Calorimeter - Editor Jessop

## 15 Pages

### 5.1 Introduction

Editors: jessop

- Cover full detector, not just US scope

Test **DETAILED OUTLINE FOR JULY 1st**

- very brief description of motivation and scope of upgrade (first six layers of HB and back-end electronics) to introduce the upgrade.
- Description of post phase 1 upgrade HCAL barrel active elements and absorber. Figure to show geometry. Describe megatiles, WLS, ODU's, SiPM's and FEE. Figure to show megatile and WLS.
- Description of radiation damage mechanism, rate dependence, effects and extrapolation of light yield estimate to 3000 fb<sup>-1</sup>
- Description of post phase 1 back-end electronics and HB trigger
- Description of trigger rate limitation

### 5.2 Requirements

Editors: Hill, Jessop

**DETAILED OUTLINE FOR JULY 1st**

- Restatement of phase 1 requirements for HCAL performance and desire to maintain phase 1 performance through phase 2 (e.g Jet reco efficiency, resolution, e/h for electrons/photons, MET)
- Trigger rate requirements HB objects

### 5.3 Proposed design

Editors: Belloni

**DETAILED OUTLINE FOR JULY 1st**

- Description of baseline new scintillator choice (double doped EJ260 with O2 fiber) and radiation and performance tests.

- Description of possible new WLS fiber configurations in megatile
- Description of back-end electronics using EB developed cards that will be adapted
- Description of hardware possibility to combine EB/HB in back-end crates for combined primitive

The CMS Hadronic Barrel calorimeter covers the pseudo-rapidity range  $0 < |\eta| < 1.3$ , and extends radially between 1.806 and 2.95 meters. It is a sampling calorimeter; it uses scintillator/wavelength-shifting fibers as the active detectors, and brass as the passive material. The wavelength-shifted scintillation light is read by silicon photo-multiplier (SiPM); their generated pulse is integrated by a QIE ASIC developed for the CMS HCAL upgrade.

The HB consists of a large number of  $\Delta\phi \times \Delta\eta$  projective towers, grouped in four samplings in depth. Each tower is  $5^\circ$  in  $\phi$ , for a total of 72 sections. The HB is divided into 18 mechanical wedges, and in each wedge a single assembly of scintillator tiles, optically divided in  $\Delta\phi \times \Delta\eta = 5^\circ \times 0.087$  units, covers a  $\Delta\phi = 20^\circ$  and  $0 < \eta < 1.3$  area. These scintillator assemblies are called trays and constitute the construction unit of the HB. Each tray is divided in three areas: left-side, central, and right-side, covering a  $\phi$  interval of  $5^\circ$ ,  $10^\circ$ , and  $5^\circ$ , respectively. The length of the trays varies between 3.7 cm and 4.3 cm; the central tray widths vary between 33 cm and 50 cm, while the side tray widths vary between 16 cm and 24 cm. The size of the individual tiles vary between  $15 \times 17 \text{ cm}^2$  and  $16 \times 40 \text{ cm}^2$ .

Each scintillator tray is 4 mm thick, with the exception of the first layer, which is 9 mm thick.

### 5.3.1 Scintillator choice

The proposed upgrade consists of a replacement of the inner six layers of the HB with a more radiation-tolerant scintillator. The baseline material identified by on-going R&D is an over-doped scintillator based on the green emitting EJ-260. This new scintillator represents an improvement in radiation-damage tolerance over SCSN-81, currently used in the HB, in three steps:

- the base material of EJ-260 is polyvinyltoluene, while SCSN-81 uses polystyrene. R&D demonstrates that PVT is less damaged by radiation;
- EJ-260 emits green light, which is less affected by the color centers created in plastic scintillator by radiation. The expression that scintillator experts adopt to describe this process is "better red than dead", indicating that emission at a longer wavelength is less susceptible to absorption by color centers in the base material;
- over-doping the primary dopant of a scintillator is shown by R&D to improve the light-yield during irradiations. The concentration of dopants is tuned in commercial scintillators to optimize the light yield before any irradiation. Increasing the dopant concentration has thus the effect to reduce the light yield of the undamaged scintillator. However, as the scintillator is irradiated, UV-absorbing radicals are created by means of splitting of the polymer chains. The presence of radicals reduces the mean free path of light, and reduces the probability that light encounters a dopant and is shifted to a longer frequency that can travel undisturbed. Increasing the dopant concentration thus increases the probability that light can be efficiently shifted in wavelength before being re-absorbed by the base material.

The usage of a green-emitting scintillator forces the change of wavelength-shifting fibers. The current HB uses Y11 green-emitting fibers produced by Kuraray; it is now necessary to purchase fibers that absorb green light and emit it at a longer wavelength. The choice is limited



to O<sub>2</sub>, red/orange-emitting fibers, also produced by Kuraray. While this configuration could have been problematic in the past, when typical photo multiplier tubes were not sensitive to long wavelength emissions, the photodetectors that will read the wavelength-shifted light are SiPM, and have a good quantum-efficiency for red/orange light.

### 5.3.2 Optical Path

The baseline design reproduces exactly the design of the current HB. Each scintillator tile, covering an area  $\Delta\phi \times \Delta\eta = 5^\circ \times 0.087$ , contains a wavelength-shifting fiber located in a  $\sigma$ -shaped groove running along the side of each tile. The WLS fiber is then spliced to a clear plastic fiber, that brings the light to the SiPM. Radiation-damage studies showed that also the Y11 green-emitting WLS fibers received a similar damage as pieces of plastic scintillator. In particular, a 2 cm by 10 cm piece of SCSN-81 scintillator, with a straight Y11 fiber installed in the center of it, reduced its light yield by a factor of 26 after an irradiation in Co-60  $\gamma$ -rays, to an integrated dose of 30 Mrad. It was shown that the light-loss is a factor of 4 due to the WLS fiber, and a factor of 6 due to the plastic scintillator.

Therefore, it is envisioned to test whether there is a need to worry about the damage incurred by the WLS fibers. An alternative design proposes to divide the scintillator cells into strips, and serve each strip with a separate WLS fiber. While this procedure does not address the issue of radiation-damage on the fiber, as soon as the fibers do not reduce their light yield to 0, any increase in light output (achieved, in this design, by reducing the light path between the production point and the WLS fiber) helps increasing the radiation tolerance of the detector.

### 5.3.3 Back-end Electronics

The baseline design consists in replacing the back-end electronics system (MCH, AMC13, uHTR, GLIB) with a system based on the new cards that is being developed for usage in the electromagnetic barrel calorimeter (EB). The new cards will offer the possibility of running at a higher bandwidth, which is necessary to bring HB safely within the requirements for HL-LHC conditions.

The DAQ requirements for both EB and HB are similar. It is expected that the same hardware will fit the requirements of both systems, with at most the need to adapt the firmware.

The usage of the same hardware opens the opportunity of combining the HB and EB outputs in the back-end crates, and, for example, provide a combined EB+HB primitive to the trigger and DAQ systems.

## 5.4 Performance

Editors: Jessop

**DETAILED OUTLINE FOR JULY 1st**

- Physics performance w/wo upgrade for scintillator/WLS choice
- Trigger performance w/wo upgrade

## 5.5 Alternatives

Editor(s): Belloni

**DETAILED OUTLINE FOR JULY 1st**

Table 5.1: Characteristics of plastic scintillators and WLS fibers considered for usage in the HB upgrade; OD indicates over-doped scintillators.

	EJ-200	EJ-260	EJ-200 OD	EJ-260 OD
Light output [% anthracene]	64	60		
Scintillation efficiency [photons/1MeV $e^-1$ ]	10,000	9,200		
Maximum emission wavelength [nm]	425	490		
Light attenuation length [cm]	380	350		
Rise time [ns]	0.9	–		
Decay time [ns]	2.1	9.2		
Polymer base	PVT	PVT	PVT	PVT
	Y11	O2		
Color	green	orange		
Peak [nm]	476	550		
Absorption peak [nm]	430	535		
Attenuation length [m]	> 3.5	> 1.5		

- Alternative Scintillator
- Alternative WLS (update: this goes with the scintillator)
- Alternative segmentation
- Alternative solution to electronics

Alternative design ideas are being investigated to guarantee the successfulness of the HB upgrade project. These alternatives include the usage of different active material, the definition of a different detector segmentation, and a different electronic setup.

### 5.5.1 Scintillator and WLS alternatives

The R&D effort devoted to testing the radiation tolerant of plastic scintillators included a comprehensive set of samples, among which the over-doped version of the green scintillator EJ-260 has been chosen as the baseline material for the upgrade. An active material more similar to the currently used scintillator (SCSN-81) is EJ-200, a PVT-based blue scintillator. Its radiation tolerance is similar to over-doped EJ-260, and this material is considered as a viable alternative to EJ-260, in case it performed similarly. Table 5.1 reports a comparison among the salient features of EJ-260 and EJ-200, and their over-doped versions.

In general, the light yield of blue scintillators is higher than green scintillators; over-doping reduces the light yield of un-irradiated scintillators (while, however, increasing their radiation tolerance); green WLS fibers have a longer attenuation length of orange ones. If EJ-200 were demonstrated to be enough radiation tolerant to provide enough light at the end of the HL-LHC period, with a large enough safety factor, it could be a strong contender. From the costing point of view, EJ-200 is about 10% cheaper than over-doped EJ-260. This alternative provides an opportunity for a cost saving.

### 5.5.2 Segmentation

The original design of the HB scintillator tiles defines a detector with the following segmentation characteristics:

- Transverse segmentation:  $\Delta\phi \times \Delta\eta = 5^\circ \times 0.087$  towers, constituted by rectangular

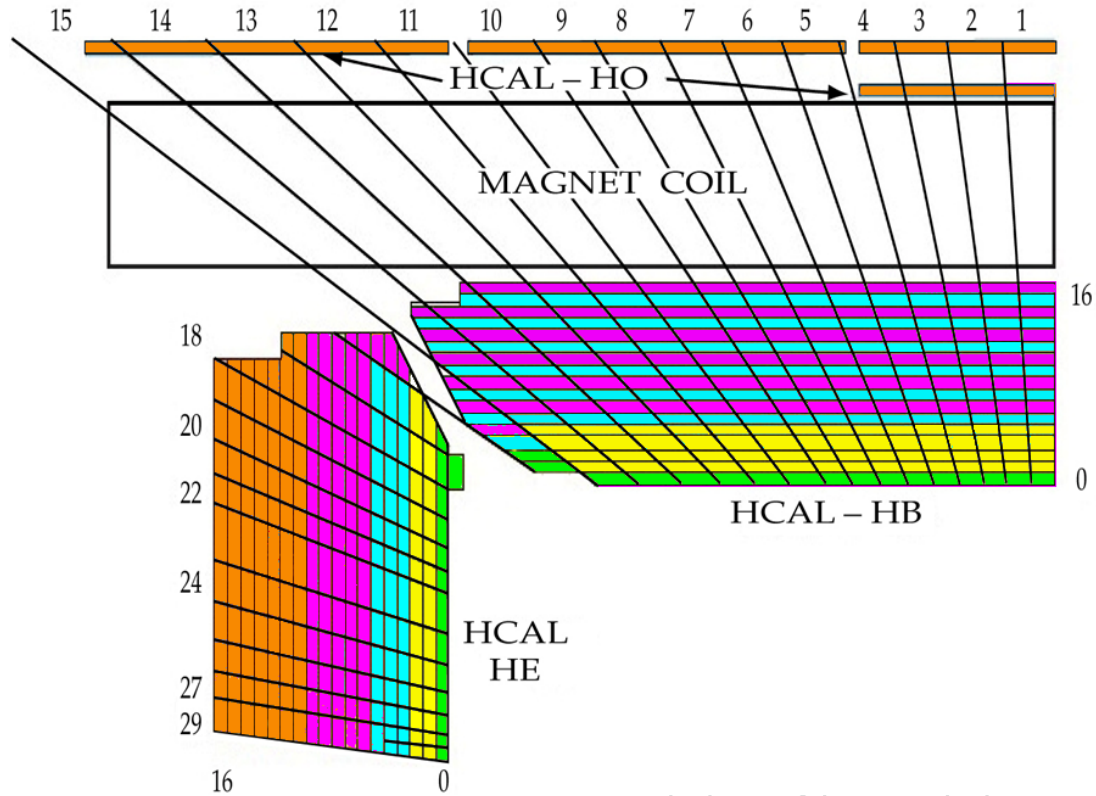


Figure 5.1: Longitudinal segmentation of the HB Phase-1 upgrade detector. Each color indicates the four sets of layers individually read [33].

tiles the size of which varies between  $15 \times 17 \text{ cm}^2$  and  $24 \times 40 \text{ cm}^2$

- Longitudinal segmentation: the tiles in the same  $\Delta\phi \times \Delta\eta$  tower are read together, with the exception of the tiles in the first layer; the longitudinal segmentation thus defines two sections (layer 0; layers 1-17)

Both transverse and longitudinal segmentations play a role in defining the performance of the detector as a function of radiation damage, and possible modifications of the current design are presented below. These modifications are being investigated as possible alternative designs.

The scintillator tiles located in the first layer are the ones that suffer the largest light-yield reduction due to radiation damage. A large component in the light reduction is due to the decrease in attenuation length, due to the creation of color centers. In order to reduce the light path from the scintillator to the WLS fiber, the proposal of cutting the square tiles, read by a single WLS fiber, into narrow finger-shaped tiles, individually read out, has been presented. The area covered by a single  $15 \times 17 \text{ cm}^2$  would be covered by five  $3 \times 17 \text{ cm}^2$  or about eight  $2 \times 17 \text{ cm}^2$ , each of which would have a single straight WLS fiber installed along the narrower tile. The WLS fibers serving the newly created finger tiles would be then merged and the sum of their light provided to a single photodetector. In this way, the number of readout channels (and hence the actual transverse segmentation) would not change, but the light output would be less affected by radiation damage in the plastic.

The HB is undergoing a Phase-1 upgrade that will result in an increase in longitudinal segmentation. In particular, the readout schema will have three longitudinal segmentations, as shown in Fig. 5.1.

An increase in longitudinal segmentation allows one to apply radiation-damage-dependent

3511 corrections with an higher granularity, thus being able to improve jet energy resolution and  
3512 response over time. However, increasing the longitudinal segmentations comes at a cost; the  
3513 number of readout channels needs to increase accordingly. In the current upgrade schema  
3514 the first six layers of the HB are being read in two separate longitudinal segments; alternative  
3515 longitudinal segmentations, which could have a significant impact on the performance of the  
3516 detector, since this is the area which will see the largest radiation damage, are being studied.

DRAFT

## Chapter 6

# Endcap Calorimeter - Editors Yohay Mans 40 pages

## 6.1 Introduction

The current CMS endcap calorimeter consists of a lead tungstate ( $\text{PbWO}_4$ ) total absorption electromagnetic section (EE), covering  $1.5 < |\eta| < 3.0$  and  $25.8 X_0$ , and a brass/scintillator sampling hadronic section (HE), covering  $1.3 < |\eta| < 3.0$ . Together, the EE and HE cover approximately  $10 \lambda_I$ . After 10 years of operation of the High Luminosity LHC (HL-LHC), corresponding to a delivered integrated luminosity of  $3000 \text{ fb}^{-1}$ , the region of the EE(HE) is expected to sustain an integrated dose up to 1.5 MGy(0.5 MGy), shown in Figure 6.1, and neutron fluences of  $10^{16}(10^{15}) \text{ n/cm}^2$ . Severe attenuation of the transmission of blue scintillation light through the EE  $\text{PbWO}_4$  crystals, caused by hadron damage to the crystal matrix, is predicted to lead to a signal loss of over 90% by the end of HL-LHC operation. Similarly, radiation induced transparency loss in the HE scintillator and optical readout fibers, coupled with changes in the scintillator dopant characteristics, is expected to lead to an even worse signal loss. Predictions of the calorimeter signal loss are shown in Figure 6.2. The expected signal loss will significantly degrade the detector's electromagnetic (EM) and hadronic energy resolution. Clearly, the active materials of the CMS endcap calorimeter must be upgraded to maintain the calorimeter's scientific utility within the context of the HL-LHC physics program (cf. Sections **[sections in the Science chapter covering WW scattering and VBF Higgs production]**).

The average number of pp collisions per 25 ns bunch crossing at the HL-LHC is expected to be 140, a factor  $\sim 3$  higher than expected for Runs 2 and 3 of the current LHC. In events with multiple pp collisions per crossing, or *pileup*, the signal in any given channel may represent the sum of signals from particles produced in separate collisions. This makes signal amplitude estimation more difficult. In addition, optimal performance of EM and hadronic clustering algorithms strongly depends on pileup, which naturally fluctuates from event to event in a given detector channel. The resulting effect of pileup in the calorimeter is to effectively worsen the EM and hadronic energy resolution, as shown in Figure 6.3. The CMS endcap calorimeter upgrade must enable efficient separation between many overlapping events within one bunch crossing, in a radiation hard construction, in order to enable the HL-LHC scientific program.

To address the unique challenges posed by the high radiation, high pileup HL-LHC collision environment, the CMS Collaboration has proposed to replace the current endcap calorimeter with a high granularity sampling calorimeter (HGCal). The primary advantages for HL-LHC operation of the proposed HGCal over the current endcap calorimeter are

- the use of silicon as active material in the areas expected to receive 50 kGy or more, which has been shown (cf. Section 6.3.3 below) to exhibit a tolerable ratio of signal



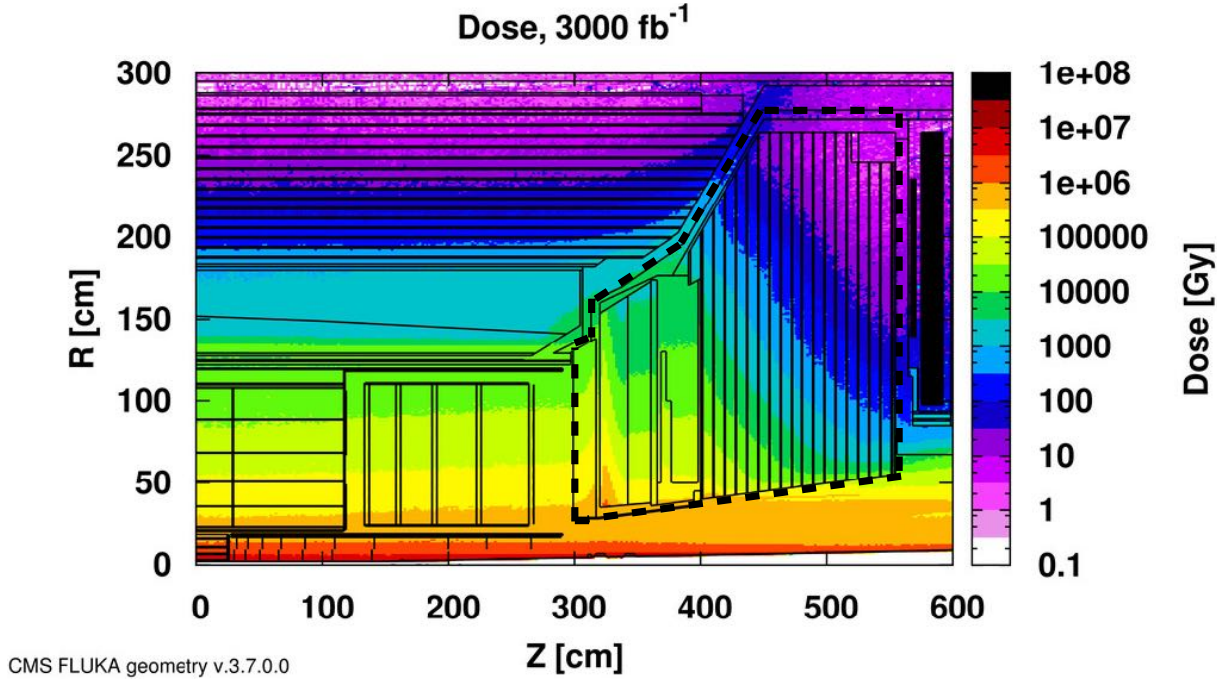


Figure 6.1: Expected integrated dose in the region of the current CMS calorimeters after 3000 fb<sup>-1</sup> of delivered luminosity. The endcap calorimeters are within the dashed black line.

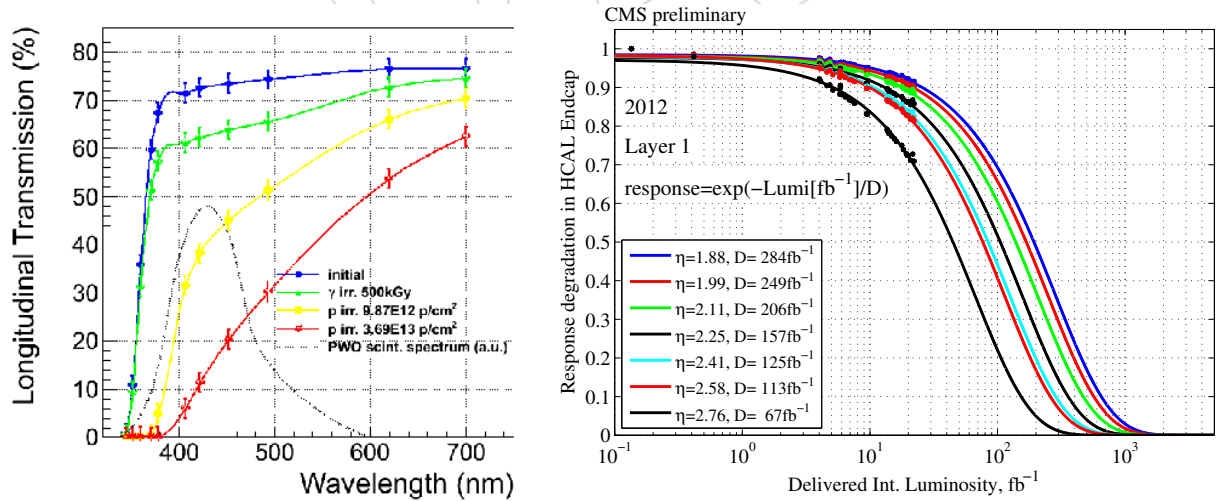


Figure 6.2: Signal loss predictions for the current endcap calorimeter active materials as a function of irradiation dose. (Left) PbWO<sub>4</sub> light transmission spectra for different  $\gamma$  and proton doses. All proton doses are lower than the maximum expected by the end of HL-LHC operation. (Right) Fractional decrease of scintillator light in HE layer 1 (closest to the interaction point) as a function of LHC integrated luminosity.

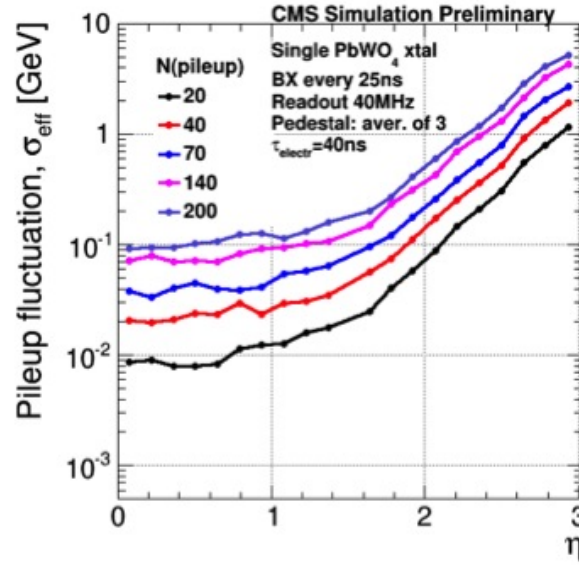


Figure 6.3: Contribution to EM energy resolution in terms of an effective noise term as a function of  $\eta$  for different  $N(\text{pileup})$  (= number of overlapping pp interactions) scenarios.

to noise (S/N) after exposure to that dose; and

- the increased channel count and number of layers, which allow particles produced in the same EM or hadronic shower to be better associated and discriminated from pileup.

As in the current calorimeter, the proposed design incorporates a front electromagnetic section (EE), covering  $26 X_0$  and  $1.6 \lambda_I$ , followed by a hadronic section (HEC) covering another  $9.1 \lambda_I$ . The 28-layer EE consists of alternating layers of silicon and copper-tungsten (CuW) absorber. The 24-layer HEC uses stainless steel as the absorber and utilizes a mixture of silicon and plastic scintillator active material, with silicon covering the high dose, high fluence regions and plastic scintillator covering the rest. Because it will exhibit less radiation induced aging and will better discriminate between pp interactions, thereby counteracting the main drivers of poor resolution, this detector design is expected to yield EM and hadronic energy resolutions on par with those of the current endcap calorimeter (cf. Section 6.5.1).

The following list summarizes the US deliverables for the new endcap calorimeter (EC).

- Silicon modules
  - Purchase of 19% of silicon
  - Construction of all HEC silicon modules
  - Construction labor for all EE odd size modules (cf. Section 6.3.3)
  - Development of silicon module construction procedures
  - Contributions to design of module printed circuit boards (PCBs)
  - Construction of module PCBs for all US-built modules
- Scintillator modules
  - Development of silicon photomultiplier (SiPM) readout technology (cf. Section 6.3.4)
  - Purchase of SiPMs
- Mechanics

- Construction of cassettes for the front 12 HEC layers (cf. Section 6.3.5)
- Development of cassette construction procedures for the front 12 (primarily silicon) and rear 12 (primarily scintillator) HEC layers
- Electronics and services
  - Design of concentrator chip (cf. Section 6.3.6)
  - Construction of motherboards for silicon modules for the US-built cassettes (cf. Section 6.3.5)
  - Design and construction of motherboards for scintillator modules for the US-built cassettes
  - Design and construction of DAQ and control electronics for the complete endcap calorimeter
  - Contributions to the development of trigger firmware

The remainder of this chapter is organized as follows. Section 6.2 connects the science requirements of the CMS Phase 2 endcap calorimeter to the specific design choices made for the EC. Section 6.3 describes the overarching EC design. Section 6.4 details the specific US deliverables listed above. Section 6.5 shows the expected performance of the EC. Sections 6.6 and 6.7 describe design alternatives that were considered and rejected and ongoing R&D to improve the design, respectively. Sections 6.8 and 6.9 respectively present an estimate of the project risks and the project value engineering. Finally, quality assurance (QA) and quality control (QC) procedures are laid out in Section 6.10, and an assessment of project environmental impact, health, and safety is made in Section 6.11.

## 6.2 Requirements

The CMS Phase 2 forward calorimeter will extend electron, photon, jet, and  $E_T^{\text{miss}}$  reconstruction to  $|\eta| = 3.0$ . Maintaining good resolution for these objects in the forward region, as well as high availability and precise calibration of the calorimeter, is a key ingredient in the following areas of the HL-LHC scientific program:

- Measurement of VBF Higgs production, a process identified by two characteristic forward jets
- Measurement of WW scattering, a process identified by two characteristic forward jets
- Electroweak physics measurements and new physics searches, as extending these analyses from  $|\eta| = 1.5$  to  $|\eta| = 3.0$  more efficiently exploits the HL-LHC dataset

Jet energy resolution at the trigger and reconstruction levels plays an important role in efficiently tagging forward jets characteristic of VBF Higgs and WW scattering. Figure 3.1 shows the projected uncertainty on Higgs couplings to different SM particles achievable with the 3000  $\text{fb}^{-1}$  HL-LHC dataset, assuming the same detector performance as in Run 1. Measurement of Higgs couplings to this precision is a baseline scientific goal of the HL-LHC physics program, and imposes a requirement on the HGCal jet energy resolution that must be fulfilled by the HGCal detector design.

Jet energy resolutions for pileup-corrected anti- $k_T$  jets with  $R$  parameter 0.5, reconstructed with the particle flow algorithm, were measured to be 5-10% for jet  $p_T > 60$  GeV and  $|\eta| < 0.5$  [77]. Published results for  $|\eta| > 1.5$  are not available, but it is expected that resolutions in the forward region of the calorimeter should not be too much worse. Furthermore, the details of jet

clustering and corrections will be different for the CMS Phase 2 calorimeter than for the Run 1 detector, but a benchmark jet energy resolution of better than 10% is still a sound technical requirement on the HGCAL in view of forward jet tagging needs at the HL-LHC.

The requirement of better than 10%  $p_T$  resolution on jets in the HGCAL drives the number of layers required in the HEC, which in turn drives the project cost. Figure 6.4 shows an estimate of the  $p_T$  resolution for true forward jets in a VBF Higgs Monte Carlo (MC) simulation as a function of number of layers in the HEC, assuming ideal calibration. The required resolution can only be achieved with a 24-layer HEC, which is a cornerstone of the HGCAL proposal.

Placeholder

Figure 6.4: Estimate of the  $p_T$  resolution for true forward jets in a VBF Higgs Monte Carlo (MC) simulation as a function of number of layers in the HEC.

Realization of the required jet energy resolution depends on precise intercalibration of detector cells. As the energy deposit of minimum ionizing particles (MIPs) in a given thickness of silicon is well understood, the HGCAL cell sizes are chosen to provide sensitivity to single MIPs throughout the HL-LHC operation period. The design cell sizes of 0.5-1 cm<sup>2</sup> are chosen to limit the cell capacitance to 60 pF, and are small enough that a MIP signal can be seen above the capacitive and leakage current noise even after the sensor charge collection efficiency has fallen due to radiation damage. MIP reconstruction can therefore be exploited to provide continuous recalibration of the detector.

Single MIP signals are expected to be visible after 3000 fb<sup>-1</sup> of delivered integrated luminosity with a signal-to-noise ratio (S/N) of 1.7-6.5, depending on location within the HGCAL. For redundancy, dedicated low-capacitance/low-noise cells will be included on each wafer, each with the ability to resolve single MIPs with S/N > 5 even after integrating 3000 fb<sup>-1</sup>. More de-

tails about module design and calibration can be found in Sections 6.3.3 and 6.5.2, respectively.

The need for single MIP sensitivity is motivated by the maximum tolerable intercell miscalibration in VBF and VBS jets in order to maintain an energy resolution of 10% or better. Figure 6.5 shows an estimate of the  $p_T$  resolution for true forward jets in a VBF Higgs MC simulation for different values of intercell calibration precision, assuming a cell size of  $\sim 1 \text{ cm}^2$  (corresponding to an 8'' module with 192 cells, cf. Sec. 6.3.3). A calibration precision of  $\mathbf{N}\%$  is applied as a smearing of the cell response with Gaussian width  $\sigma_E/E = \mathbf{N}\%$ . An intercalibration precision of  $\mathbf{M}\%$  is required to meet the jet energy resolution goal of 10% for  $p_T > \mathbf{20} \text{ GeV}$ . With a worst case S/N of 1.7, implying a noise level of 0.6 MIP expected in the regions of highest dose after  $3000 \text{ fb}^{-1}$  of integrated luminosity, such a precision can be met with a sample of  $\mathbf{P}$  million Level-1 accepts (L1As).

Placeholder

Figure 6.5: Estimate of the  $p_T$  resolution for true forward jets in a VBF Higgs MC simulation for different values of intercell calibration precision, assuming a cell size of  $\sim 1 \text{ cm}^2$ .

Figure 6.6 shows the expected per-cell noise increase as a function of fluence for different cell sizes. The noise can be maintained below 0.6 MIP in the regions of highest fluence when a cell size of  $\sim 0.5 \text{ cm}^2$  is used, and for the rest of the HGCAL when a size of  $\sim 1 \text{ cm}^2$  is used. This insures a sufficiently precise MIP intercalibration to enable high resolution VBF and VBS jet energy measurements.

Resolution for EM objects is also a design driver for the HGCAL. To contribute a maximum of 0.5% to the EM energy resolution constant term, an intercalibration precision of 3% must be maintained in the HGCAL EE. In the worst case fluence scenario, this can be met using MIP tracking in  $0.5 \text{ cm}^2$  cells with 150 million L1As. With this resolution, endcap performance is



Placeholder

Figure 6.6: Expected per-cell noise increase as a function of fluence for different cell sizes.

not very different from barrel calorimeter performance, and diphoton, dielectron, and  $E_T^{\text{miss}}$  resolutions are expected to be similar to that of the current detector.

The proposed HGCAL longitudinal and transverse segmentation is a direct consequence of the requirement of sensitivity to jets at high  $\eta$  that provide a telltale signature of VBF Higgs production or  $WW$  scattering. Sensitivity is maintained with a 24-layer HEC allowing high energy showers to be fully sampled, as well as small cell sizes of  $0.5\text{-}1\text{ cm}^2$  that enable MIP intercalibration to be exploited throughout the HL-LHC run period.

## 6.3 Proposed design

### 6.3.1 Design overview

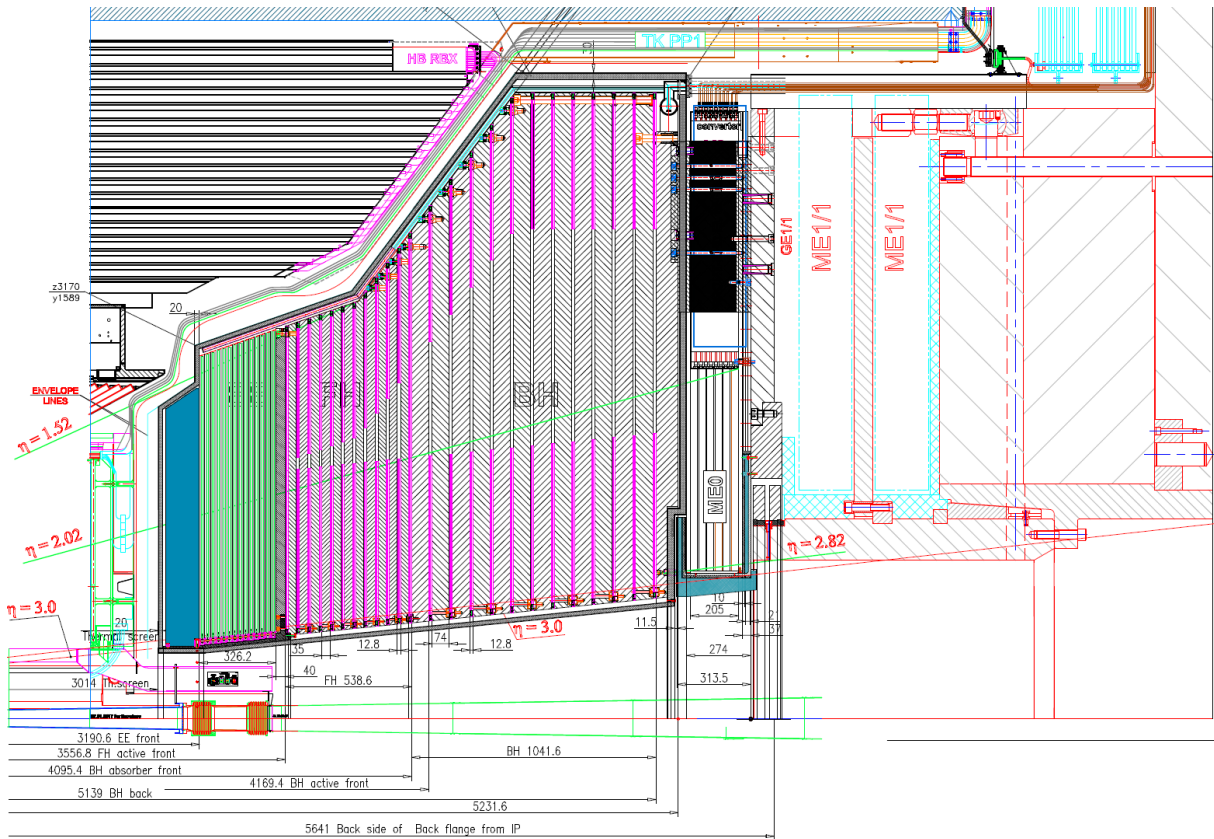


Figure 6.7: Quarter longitudinal cross section of the HGCAL.

A mechanical drawing of the proposed Endcap Calorimeter is shown in Figure 6.7. The electromagnetic section (EE) of the HGCAL extends from  $|z| = 3.2\text{ m}$  to  $|z| = 3.5\text{ m}$ , covering a depth of approximately  $26 X_0$  and  $1.6 \lambda$  with 28 sampling layers. The hadronic calorimeter, labeled FH and BH in Fig. 6.7, extends  $|z| = 3.5\text{ m}$  to  $|z| = 5.1\text{ m}$ , covering a depth of approximately  $9.1 \lambda$ . The hadronic section is divided into finer and coarser sampling section, each comprising 12 layers. The two sections extend from  $|z| = 3.5\text{ m}$  to  $|z| = 4.1\text{ m}$  and  $|z| = 4.1\text{ m}$  to  $|z| = 5.1\text{ m}$  and cover depths of  $3.2 \lambda$  and  $5.9 \lambda$  respectively.

The EE is made of alternating absorber and silicon detector layers. The absorber in turn is made of alternating layers of stainless steel-clad lead and sandwiches of copper-tungsten and copper. The silicon sensors are mounted on the two sides of a 6 mm thick copper plate which is both part of the absorber and a cooling plate in which a small diameter stainless steel tube carries

liquid CO<sub>2</sub>. The cooling plate is formed as a wedge extending over the full radius of each layer and subtending an azimuthal angle of 60°. The silicon detectors are built in hexagonal modules with a corner-to-corner dimension of 190 mm. These are mounted on baseplates made of Cu-W of a thickness chosen to provide, together with the copper cooling plate, one absorber sample. The other absorber layer is provided by layers of stainless steel-clad lead of the same dimensions as the cooling plate attached above and below the sensor layers. Front-end ASICs are mounted directly on each module, and groups of six modules are serviced by a “motherboard.” This assembly of lead, active detectors, copper-tungsten, copper, active detectors, lead is called a cassette. The cassettes are more fully described in Section 6.3.2, the silicon modules in Section 6.3.3, and the electronics that reads out the silicon sensors in Section 6.3.6.

The hadron calorimeter (HE) is similarly made of alternating absorber and detector layers. The absorber is made of disks of stainless steel of 35 mm thickness for the first 12 samples and 74 mm for the last 12. The active detector layers are made of silicon detectors similar to those in the EE in the first 8 layers. In subsequent layers silicon detectors are used in the smaller radius (higher  $\eta$ ) regions and scintillator, read out by silicon photomultipliers (SiPM) mounted on the scintillator tiles, at larger radius (lower  $\eta$ ). The scintillator detector modules are described in Section 6.3.4. The division between scintillator and silicon detectors is determined mainly by the radiation environment. Scintillator tiles plus SiPMs are used where the expected integrated radiation dose and hadron fluence are low enough that these detectors will survive for the full HL-LHC integrated luminosity of 3000 fb<sup>-1</sup> and silicon is used in the higher radiation regions. Maps of the expected dose and hadron fluence after 3000 fb<sup>-1</sup> HL-LHC integrated luminosity are shown in Figure 6.8, which also indicates the radial boundary between the scintillator and silicon detector regions. The radius at which the transition from silicon to scintillator is made varies with depth from about 1.5 m ( $\eta = 1.7$ ) at layer 9, to about 0.9 m ( $\eta = 2.4$ ) in the back-most layers. The silicon and scintillator sensors are assembled into cassettes similar to those for the EE, but with sensors on only one side of the cooling plate, baseplates of a less-expensive material than Cu-W (chosen for proper mechanical properties), and simple copper or aluminum covers rather than lead. The all-silicon HE cassettes subtend an azimuthal angle of 60° as in the EE. The mixed cassettes, due to their greater radial size, are built in 30° units to keep them of manageable size.

Here we need a paragraph with overview of electronics, DAQ and trigger:

- Front-end electronics: silicon module PCB, Si motherboard with concentrator, scintillator PCB and motherboard
- DAQ architecture and hardware
- trigger architecture and algorithms

### 6.3.2 Cassettes

The sensors in each layer of the EC are grouped into 30° or 60° wedges called cassettes. These are the major detector subassembly of the EC, which are subsequently assembled into full disks in the EE or inserted between absorber layers to form full disks of detectors in the HE. In the electromagnetic section, the absorber layers are part of the cassette and the EE consists of the stack of cassettes, supported by a full-disk back plate and an inner cone at the  $|\eta| = 3$  boundary. In the hadronic section, the absorbers and cassettes are independent one of the other.

Figure 6.9 is a drawing of an EE cassette. It is built by tiling the hexagonal silicon modules on both sides of the central cooling plate, which provides the main mechanical structure of the cassette. The corners of the hexagons are slightly flattened (“mouse bites”), which provides a

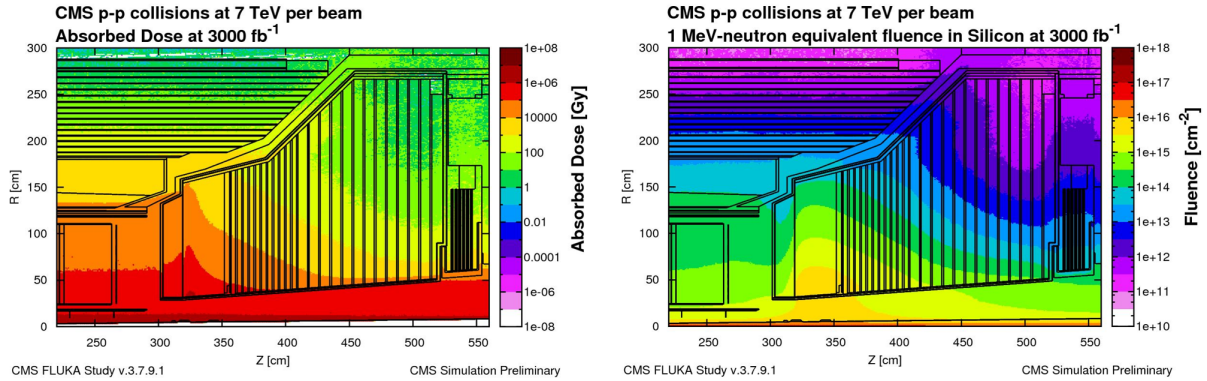


Figure 6.8: Predicted radiation dose and hadron fluence in the HGCAL after  $3000 \text{ fb}^{-1}$  HL-LHC integrated luminosity. The radius as a function of depth at which the transition is made from silicon to scintillator detector is indicated. **Draw in the boundary.**

triangular gap wherever three modules meet, as shown in the upper-right inset in the figure. A pair of screws, one from each side, secure the corners of the three adjacent modules. The lower-right inset in Fig. 6.9 is a cross-section through the thickness of the cassette at the point where a screw secures the modules showing: the cooling plate in the middle; modules made of WCu baseplates, Si sensors and PCBs on both sides; and the two covers. The covers, which also serve as half-absorber layers, consist of lead sheets clad with thin layers of stainless steel. The side facing the detector is copper coated to provide a high-conductivity grounding/shielding layer. Plates in the upper-left and upper-right corners are used to connect one cassette to the next as the cassettes are formed into a complete disk. Each cassette is also attached to the central cone (see Section refsec:Mechanical Design below) via a hole in the cooling plate and an interlocking tab on the cone.

The cassette in Fig. 6.9 corresponds to the 14th (last) cassette layer, which consists of detector layers 27 and 28. It measures approximately 1.3 m in the radial direction by 1.6 m wide across the outer corner. More upstream cassettes are, of course, smaller. The cassettes in the first disk (detector layers 1 and 2) is about 0.1 m smaller in each dimension. In moving from layer to layer, although the inner and outer radii of the cassettes change, the radial position of each “row” of modules remains the same, constrained by the fixed tiling of the fixed size hexagons.

A 4 mm diameter thin-wall stainless steel tube embedded in a serpentine groove in the cooling plate carries high-pressure two-phase  $\text{CO}_2$  fluid to maintain the detector at a temperature of  $-30^\circ\text{C}$ . This is illustrated in Fig. 6.10. This figure is for a simple  $30^\circ$  wedge-shaped cassette as used for the mixed silicon/scintillator cassettes (see below). The first 600 mm of the cooling tube, from the entrance to the bend marked “B” in Fig. 6.10 is a reduced inner diameter capillary which serves to define the main hydraulic impedance, so as to equalize the flow among different size cooling plates with different heat loads.

Cassettes with silicon modules are used in the first eight layers of the hadron calorimeter. They are of very similar design to those in the EE, with the following differences: the cassettes are single-sided with modules mounted on only one side of the cooling plate; the cover serves only as a mechanically and electrically protective layer and is therefore made of 0.5 mm thick copper; and the cassettes are individually connected to the adjacent steel absorber layer rather than to each other.

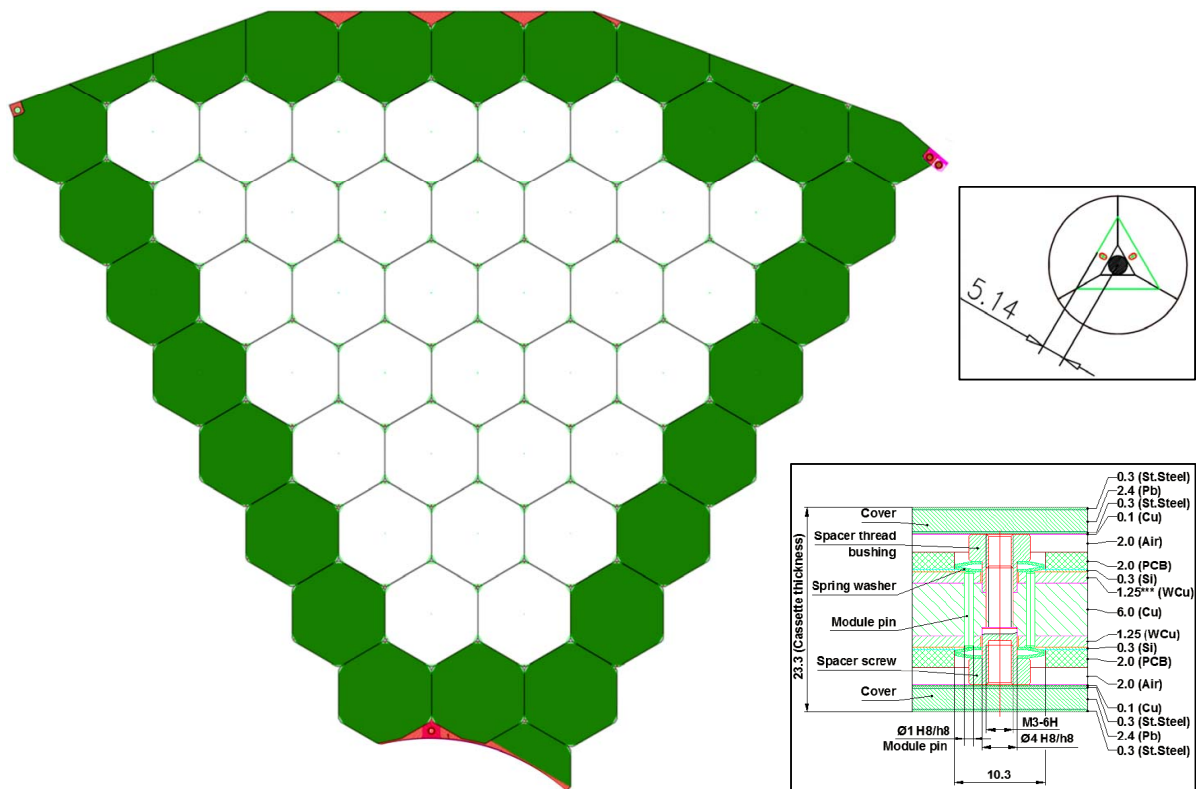


Figure 6.9: Drawing of an EE cassette. The cassette shown is for the 14th (last) cassette layer, which includes detector layers 27 and 28. This cassette is approximately 1.3 m long in the radial direction by 1.6 m wide across the outer corners. The upper right inset shows the corner where three modules meet and are secured by a common screw and spring washer. The lower right inset is a cross-section through the thickness of the cassette at the position where the screw secures the modules. **Need to update cross-section to be one with motherboards.**



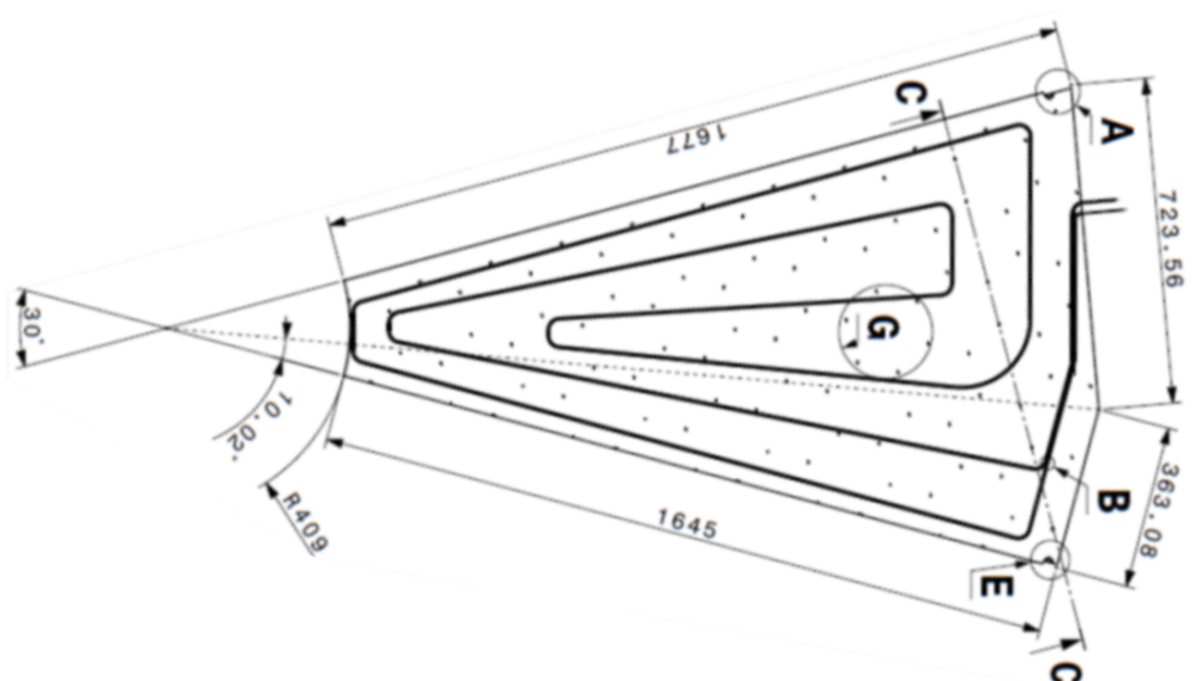


Figure 6.10: Cassette cooling plate showing the routing of the stainless steel tube that carries the high-pressure two-phase CO<sub>2</sub> coolant. This cooling plate corresponds the layer 12 of the hadronic calorimeter. **Need a better version without the extraneous fractions of a mm.**

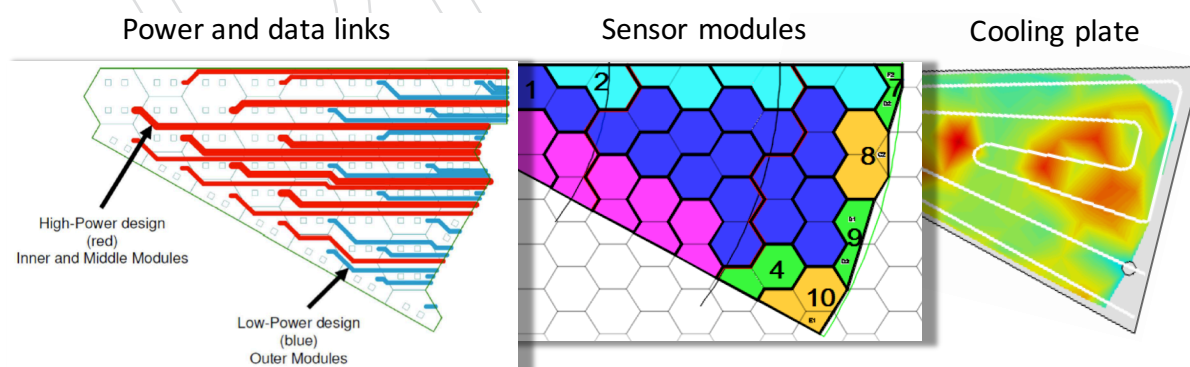


Figure 6.11: FH cassette concept. Individual modules are mounted on a copper cooling plate, and flat cables carry services from the module PCB to the outer edge of the cassette

### 6.3.3 Silicon modules

A hexagonal FH module exploded view is shown in Figure 6.12. It consists of a copper baseplate, a sensor wafer with 256 pads, and a printed circuit board with readout electronics. The sensor backplane, located on the bottom side of the sensor in Fig. 6.12, carries the high voltage bias, and is therefore insulated by a layer of Kapton foil between it and the baseplate. The top side of the sensor connects to a PCB layer, with pad connections to the PCB done with wire bonds. The front end readout chips (4 per 256 channels) are flip-chip bonded to the PCB. The hexagonal FH module is approximately 8 inches in diameter, with the sensor layer being cut from a single 8" silicon wafer. The total module thickness is approximately 3 mm.

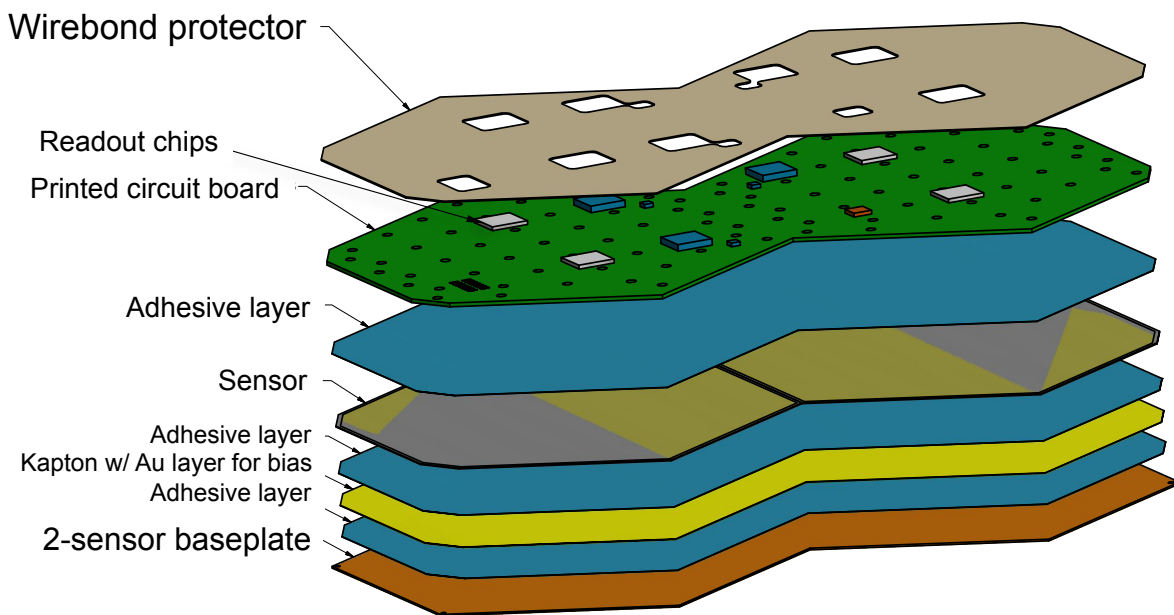


Figure 6.12: Module exploded view. **Replace with 8" diagram.**

The thermal deformation of this configuration has been tested by cooling a test setup from 23° C. to -41° C. The modules are screwed and pinned in place as they would be for installation of the cassette. The distortion is checked by measuring the capacitance between the baseplate and the copper cooling-plate. The screws and pins are then released to examine the distortion of the cold module. A maximum deflection of 0.8 mm is measured using precision shims, and this distortion is such as to press the module, when it is held by the screws and pins, against the copper cooling-plate. The resulting stress in the Si sensor is  $\approx 6$  MPa, giving a safety factor of  $> 1000$  since the ultimate tensile strength of silicon is 7 GPa.

Sensors are constructed of radiation hard n-in-p type silicon, with three different active thicknesses utilized throughout the detector. As mentioned previously, the cell area for each thickness is adjusted so as to limit the all cell capacitances to no more than 60 pF. Table 6.1 shows the regions of FH where the different active thicknesses are used, together with the maximum neutron fluence expected after 3000 fb<sup>-1</sup> of HL-LHC integrated luminosity, the cell sizes, and the expected signal-to-noise ratio ( $S/N$ ) for a minimum-ionizing particle (MIP) signal before and after 3000 fb<sup>-1</sup>. The regions are specified by the radius,  $R$ , measured from the beam axis. The regions are depicted graphically in Figure 6.14. Figure 6.13 shows that for the maximum integrated fluence expected in each region, at 600 V bias the MIP charge collected is still above  $\sim 4$  ke everywhere. The cell size and sensor active thickness are chosen such that at the end of

life of the detector, assuming a coolant temperature of  $-30^{\circ}\text{C}$ ,  $S/N$  for a single MIP is at least 1.5, large enough for the MIP signal to be calibrated.

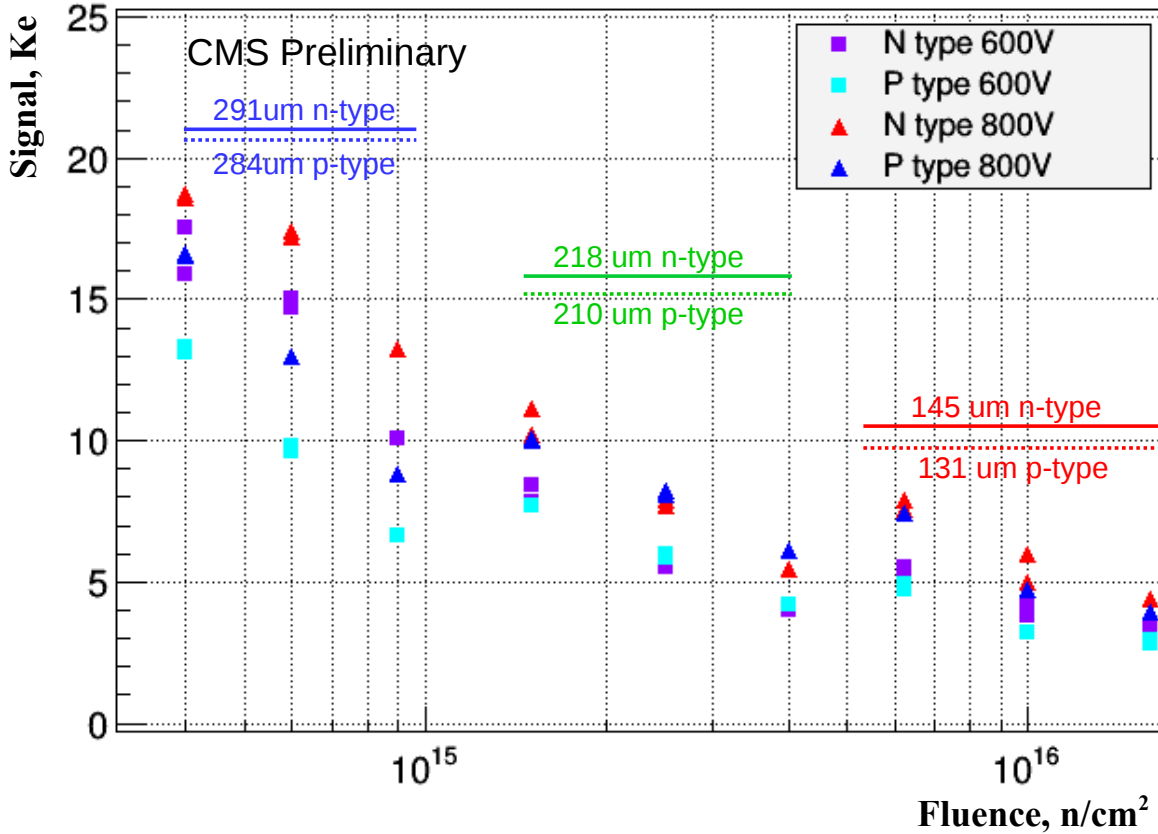


Figure 6.13: MIP charge collected vs. neutron fluence for six different types of sensors ( $\sim 300\text{ }\mu\text{m}$  n-type,  $\sim 300\text{ }\mu\text{m}$  p-type,  $\sim 200\text{ }\mu\text{m}$  n-type,  $\sim 200\text{ }\mu\text{m}$  p-type, and  $\sim 100\text{ }\mu\text{m}$  n-type,  $\sim 100\text{ }\mu\text{m}$  p-type) at two different bias voltages (square markers denote 600 V, triangle markers denote 800 V). The  $\sim 300\text{ }\mu\text{m}$  measurements extend from  $4\text{--}9 \times 10^{14}\text{ n/cm}^2$ , the  $\sim 200\text{ }\mu\text{m}$  measurements extend from  $1.5\text{--}4 \times 10^{15}\text{ n/cm}^2$ , and the  $\sim 100\text{ }\mu\text{m}$  measurements extend from  $6\text{--}15 \times 10^{15}\text{ n/cm}^2$ .

The FH module front end readout chips have 70 channels and include for each channel an amplifier, a 40 MHz low-power ADC and a TDC for signal digitization, as well as logic for digital data handling. The 70 channels comprise 63 standard sized cells plus a further cell subdivided into seven small calibration pads. The calibration pads have lower capacitance and noise than the standard pads and are designed to maintain  $S/N > 5$  throughout the lifetime of the detector, thereby providing a precise MIP calibration even after  $3000\text{ fb}^{-1}$  of HL-LHC integrated luminosity. There will be either 4 or 8 front end chips on a module, according to the number of cells. Chip fabrication is done in the 130 nm CMOS process, which has been qualified up to 400 MRad and is expected to maintain good analogue performance above the maximum dose of 150 MRad expected in the calorimeter. This technology is also known to be radiation tolerant up to very high fluences (a few  $\times 10^{15}\text{ n/cm}^2$ ), although tolerance up to  $10^{16}\text{ n/cm}^2$  will have to be demonstrated.

The front end readout chip can provide data for trigger primitive generation at the nominal bunch crossing rate of 40 MHz and will provide data readout at the nominal L1A rate of 1 MHz. Each module produces up to 6 Gb/s in Level-1 trigger data, and up to 3.2 Gb/s in full resolution data. For the modules with 100(200)  $\mu\text{m}$  sensor active thickness, 2(1) 5 Gb/s Twinax

Table 6.1: FH silicon sensor arrangement: thickness of active silicon layer with the associated cell size and  $S/N$  for a MIP before and after an integrated luminosity of  $3000 \text{ fb}^{-1}$ .

Thickness	300 $\mu\text{m}$	200 $\mu\text{m}$	100 $\mu\text{m}$
Maximum dose (Mrad)	3	20	100
Maximum n fluence ( $\text{cm}^{-2}$ )	$6 \times 10^{14}$	$2.5 \times 10^{15}$	$1 \times 10^{16}$
FH region	$R > 100 \text{ cm}$	$100 > R > 60 \text{ cm}$	$R < 60 \text{ cm}$
Si wafer area ( $\text{m}^2$ )	290	203	96
Cell size ( $\text{cm}^2$ )	1.05	1.05	0.53
Cell capacitance (pF)	40	60	60
Initial $S/N$ for MIP	13.7	7.0	3.5
$S/N$ after $3000 \text{ fb}^{-1}$	6.5	2.7	1.7

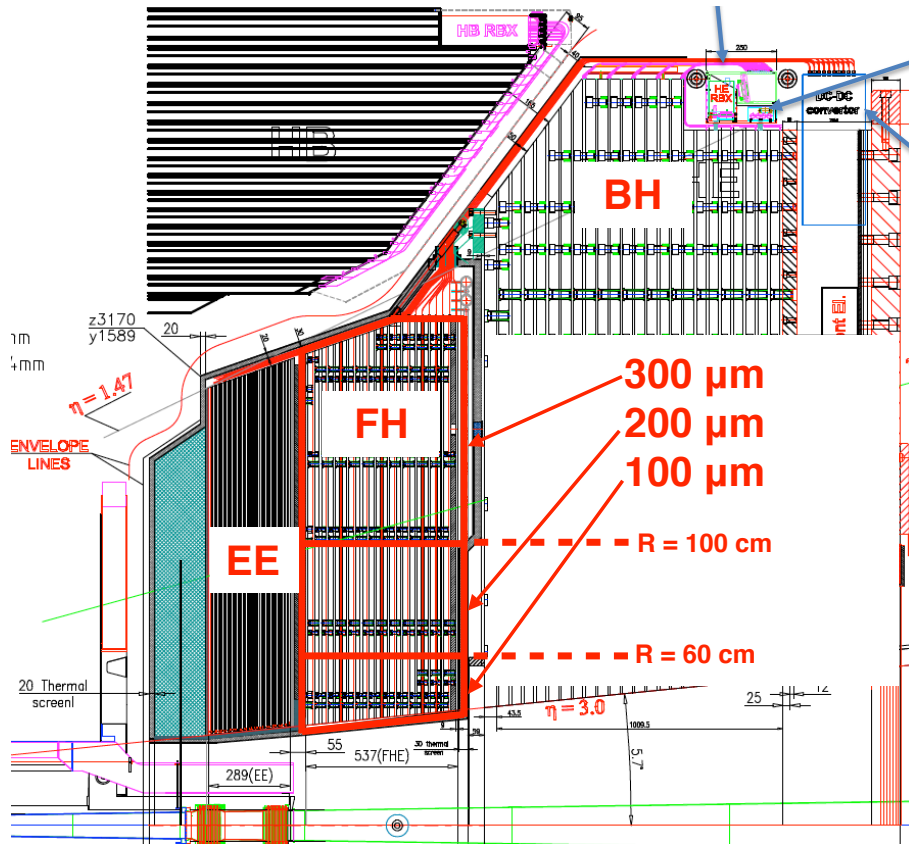


Figure 6.14: Placement of sensors by active thickness in the FH. **Replace with 1 MeV n equivalent map.**

bidirectional electrical link(s) are used to transfer full resolution data from the module to the RBX region of the endcap behind the BH, as well as carry clock and control information. These links are arranged as a ribbon in a thin, flexible support strip. The Twinax cables have a profile of less than 1 mm, are foldable and can be edge soldered to a PCB, allowing them to fit within the available space inside the calorimeter. For the 300  $\mu\text{m}$  modules, 1 10 Gb/s bidirectional optical link is shared among 4 neighboring modules for full resolution data transfer and clock and control functionality. In addition, approximately half the FH modules will be cabled to provide trigger data. The 100(200)  $\mu\text{m}$  triggering modules will carry an extra 4(2) electrical trigger links per module, while an extra 2 optical trigger links will be shared among 4 300  $\mu\text{m}$  modules.

Given space constraints and concerns about radiation hardness, DC-DC converters for regulating the module low voltage are placed in the RBX region behind BH, where the expected dose is 100 times less than at the cassette edge. A cassette low voltage PCB connects the individual FH module PCBs to the outer edge of the cassette, where flat cables bridge the connection to DC-DC converters in the RBX region. Similarly, small diameter wires carrying high voltage connect the module bias layer to a patch panel at the RBX region.

#### 6.3.4 Scintillator modules

The BH megatiles, constructed of plastic scintillator woven with WLS fibers, must withstand up to 5 Mrad of integrated dose in the regions closest to the beam pipe. Radiation hardness, ease of assembly and insertion, and ability to monitor response over time are the chief design concerns for the BH megatiles. In order to reduce the light path from scintillator to WLS fiber and thereby limit the radiation induced reduction in response, a tile geometry with a higher density of fibers per scintillator tile will be used in the highest dose regions of BH. The traditional sigma-shaped arrangement of the WLS fibers will be used in the low dose regions. The two different tile designs are shown in Figure 6.15.

In addition to fiber arrangement, a suitably radiation hard scintillator will be used for the megatile construction. Ongoing R&D to identify such a scintillator is covered in Section 6.7.

#### 6.3.5 Mechanical design

#### 6.3.6 Electronics and electrical systems

Each channel of a silicon module is wirebonded to the module PCB. Copper traces on the module PCB carry the electrical signals from each individual channel to a surface mounted custom front end ASIC (FE ASIC) serving 72 channels. 192-channel modules have 3 FE ASICs mounted, while 432-channel modules have 6. The FE ASIC performs the following functions:

- Digitization of the signal amplitude and time of arrival
- Buffering of hits from different LHC bunch crossings
- Threshold comparison
- Summation of amplitudes from neighboring channels to form inputs to the HGCal trigger primitive generator (TPG) located in the service cavern
- Transmission of TPG inputs at the LHC bunch crossing rate of 40 MHz
- Transmission of full granularity data upon receipt of an L1A at the L1A rate of 1 MHz

For signals up to  $\sim 100$  fC, the amplitude is digitized in a 10-bit ADC. For signals from 100 fC to 10 pC, the hit time is additionally digitized in an 11-bit TDC that can operate in either



## HCAL Endcap Megatiles Upgrade

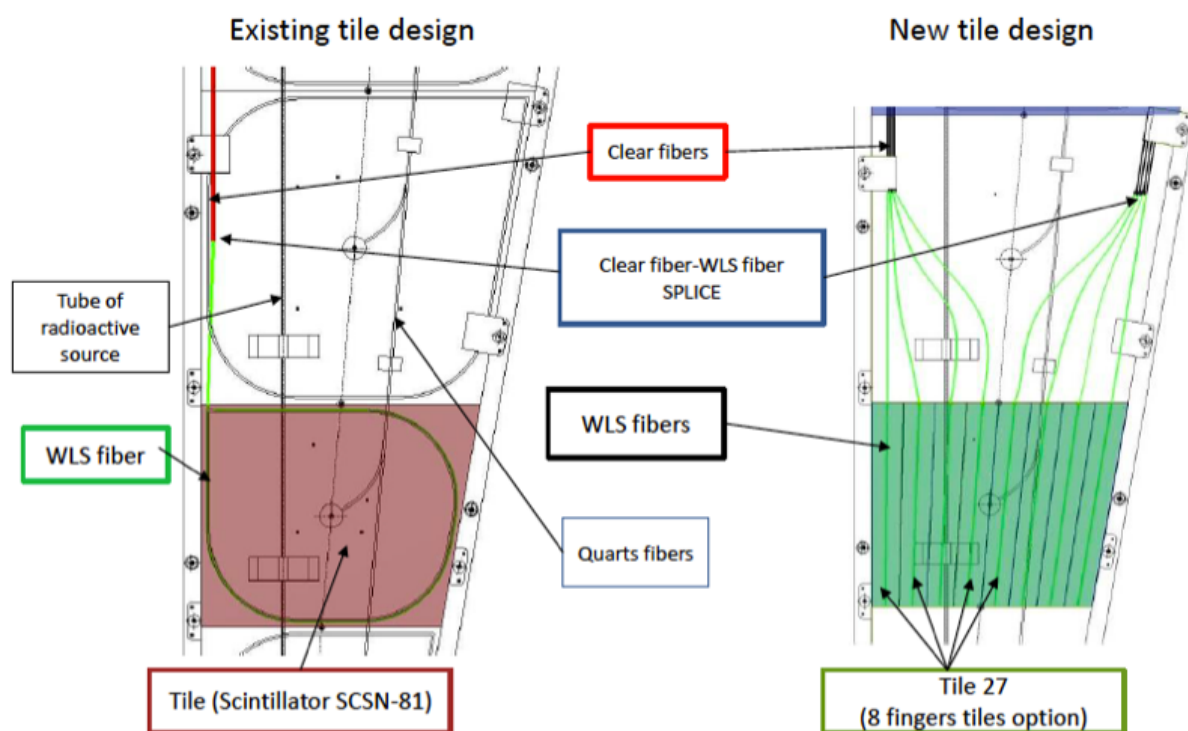


Figure 6.15: Comparison between a plastic scintillator sigma tile (left), currently used in the HE detector and proposed for the low dose regions of BH, and a finger tile (right), the proposed design for the high dose regions.

time-of-arrival or time-over-threshold mode. To recover small signals that may be affected by timewalk, hits in the previous 25 ns LHC bucket can be read out if above a high threshold.

Groups of three modules are individually connected to a single motherboard that provides the following:

- Logical grouping of modules
- Concentrator ASIC for TPG input data reduction and readout data buffering
- 1.28 Gbps electrical links for module data transfer to concentrator ASIC
- lpGBT [78] 10-Gbps bidirectional optical link drivers for clock, control, and data transmission between the modules and motherboard and the TPG and DAQ electronics
- Distribution of LV (via voltage regulators) to module FE ASIC, optical link drivers, and concentrator ASIC
- Distribution of HV to module sensor

A diagram of the motherboard-module stackup showing the mounted components and connections is given in Figure 6.16. Each motherboard has a power connector, to which LV and HV arrive from power supplies in the experimental cavern, and optical connectors for the 10-Gbps data links. Figure 6.17 shows a top (x-y) and side (z-y) view diagram of the motherboard connected to three modules.

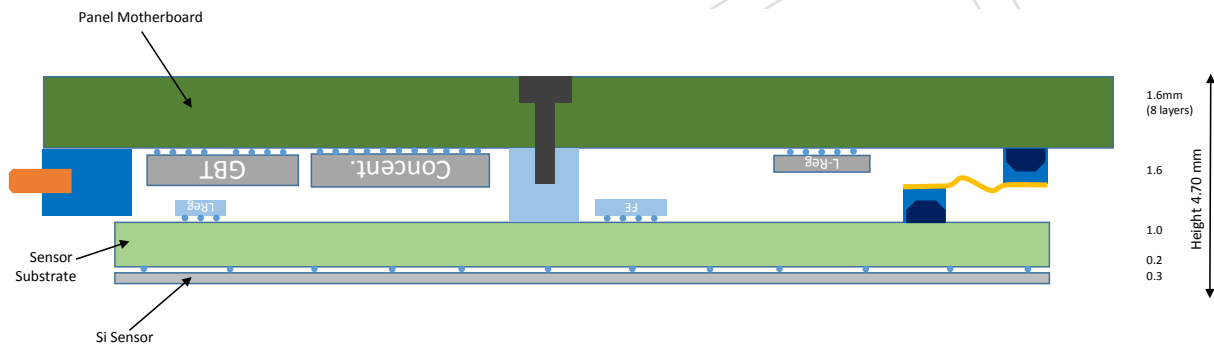


Figure 6.16: Diagram (not to scale) of the motherboard-module stackup showing the silicon sensor, module PCB (“Sensor Substrate”), FE ASIC its linear LV regulator on the module PCB, GBT 10-Gbps optical link, concentrator chip, linear LV regulator for the motherboard components, and flexible connection between motherboard and module.

Scintillator modules are read out with the same FE ASIC, concentrator ASIC, and optical links as silicon modules, but the placement of the motherboard differs. In the scintillator case, the FE ASIC is mounted to the opposite face of the readout PCB as the SiPMs, with channels milled into the copper cooling plate to allow space for the FE ASICs. The scintillator motherboards are placed at the edge of the cassette, rather than covering the modules as in the silicon case. Scintillator modules at constant  $\phi$  are ganged together and to the motherboard at the cassette edge, thereby allowing SiPM signals to pass from modules to motherboard and power and control signals to pass in the opposite direction. In the case of mixed silicon-scintillator cassettes in HEC, the space above the outer scintillator modules is used for routing power and data links to the inner silicon modules. A comparison of the silicon and scintillator PCB-module stackups is shown in Figure 6.18. A diagram of motherboard and ganged readout PCBs covering a single  $10^\circ$  wedge of HEC is shown in Figure 6.19.

Both the TPG and DAQ electronics are housed in xTCA crates in the service cavern, where they

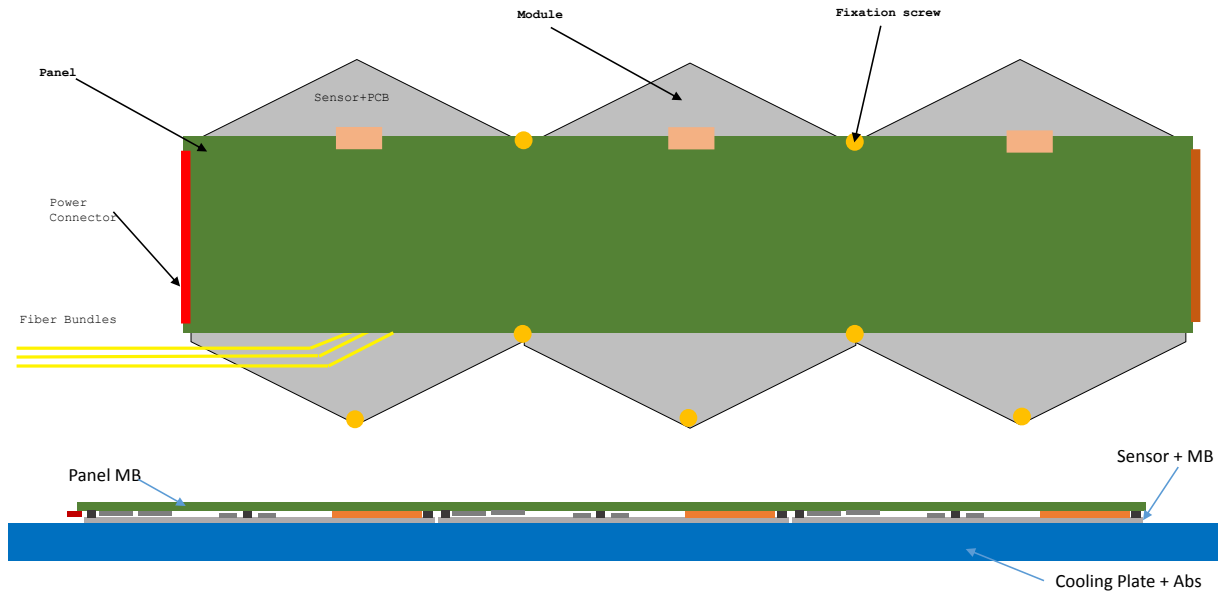


Figure 6.17: Top (x-y) and side (z-y) views of the motherboard connected to three modules.

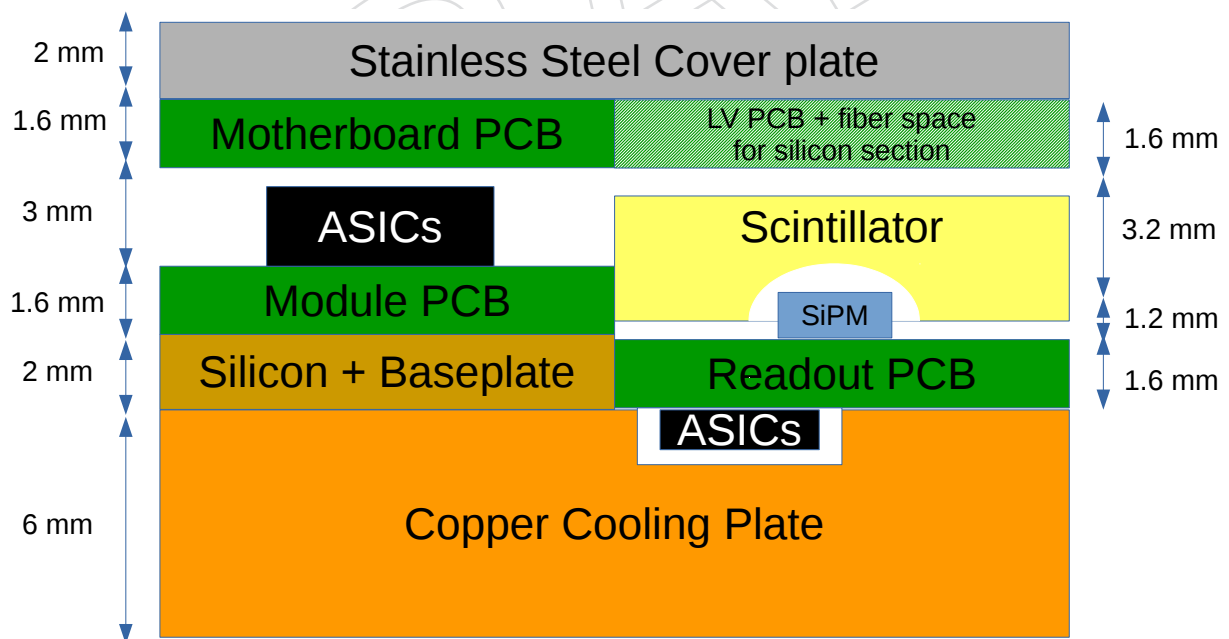


Figure 6.18: Comparison of the silicon (left) and scintillator (right) PCB-module stackups.

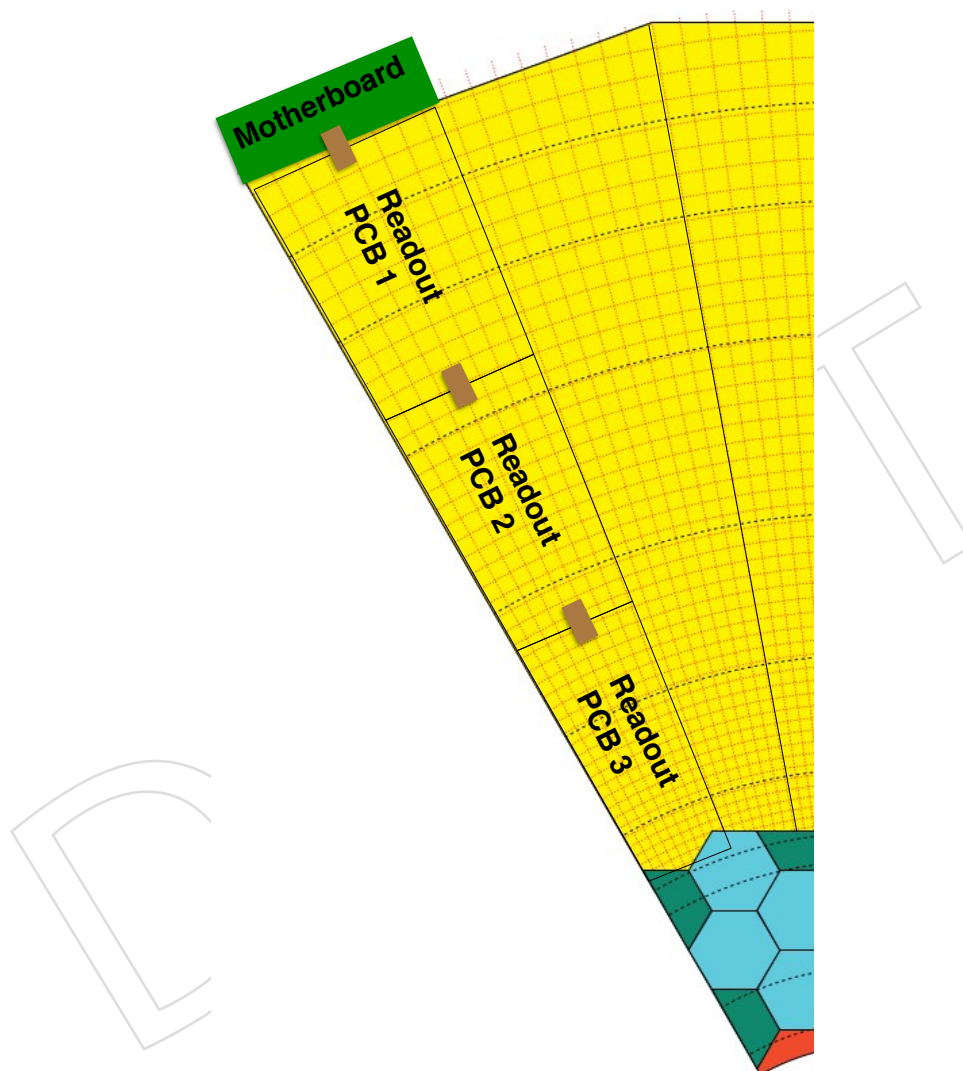


Figure 6.19: Diagram of motherboard and ganged readout PCBs covering a single  $10^\circ$  wedge of HEC.

connect to the L1 trigger and central DAQ, respectively.

### 6.3.6.1 Silicon module trigger and data readout

Assuming an average of 200 pp interactions per HL-LHC bunch crossing and a data format where each pad is read out at the L1A rate of 1 MHz with zero suppression below  $\sim 0.5$  MIP, the maximum readout data rates expected from the 300  $\mu\text{m}$ , 200  $\mu\text{m}$ , and 100  $\mu\text{m}$  modules are 580 Mb/s, 1.0 Gb/s, and 3.6 Gb/s, respectively. These rates drive the choice and number of data links described in Sec. 6.3.3, which are conservative. Full resolution front end data from the entire FH will only be transferred to the service cavern off detector electronics upon receipt of a Level-1 accept, expected at a rate of 1 MHz.

FH trigger data will be generated from sums of adjacent channels. 16 or fewer sums per 64-channel ASIC will be made with a granularity of  $2 \times 2$  sensor pads and sent at the full LHC bunch crossing rate of 40 MHz by the front end electronics to the service cavern where trigger primitives will be generated. In this scheme, the maximum trigger data rates expected from the 300  $\mu\text{m}$ , 200  $\mu\text{m}$ , and 100  $\mu\text{m}$  modules, assuming all 16 sums are transferred with no zero suppression, are 20 Gb/s, 20 Gb/s, and 40 Gb/s, respectively. To save optical link bandwidth, a concentrator chip is envisaged for each module that picks the best 12(24) trigger sums per 300  $\mu\text{m}$ /200  $\mu\text{m}$  (100  $\mu\text{m}$ ) module and sends them on to optical conversion in the RBX region (for the 200  $\mu\text{m}$  and 100  $\mu\text{m}$  modules) or directly to the service cavern via optical links (for the 300  $\mu\text{m}$  modules). It is assumed that the data from two 5-Gbps electrical links will fit into one 10-Gbps optical link. The concentrator chip reduces the maximum trigger data rates to 6.4 Gb/s, 6.4 Gb/s, and 13 Gb/s for the 300  $\mu\text{m}$ , 200  $\mu\text{m}$ , and 100  $\mu\text{m}$  modules, respectively. Table 6.2 summarizes the approximate maximum bandwidth and corresponding number of links emanating from the concentrator chip in the different FH modules. Sufficient electrical links will also be needed to transfer the trigger data on-module from the FE ASIC to the concentrator chip, but they are not detailed in Table 6.2. Studies are ongoing (cf. Section 6.7.2) to optimally allocate the links in a way that is consistent with the physics goals of the Phase II upgrade. **A diagram summarizing the TP readout scheme a la Roger's diagram would help here.**

Table 6.2: Maximum bandwidth per module for trigger and readout data as compared to current allocation of optical or electrical links emanating from the concentrator chip for the FH modules. “o-link” refers to optical links and “e-link” refers to Twinax electrical links.

Module type	Readout data bandwidth (Gb/s/module)	Readout link allocation per module	Trigger data bandwidth (Gb/s/module)	Trigger link allocation per module
FH 300 $\mu\text{m}$	0.58	$0.25 \times 10\text{-Gbps}$ o-link	6.4	$0.5 \times 10\text{-Gbps}$ o-link
FH 200 $\mu\text{m}$	1.0	$1 \times 5\text{-Gbps}$ e-link	6.4	$2 \times 5\text{-Gbps}$ e-link
FH 100 $\mu\text{m}$	3.6	$2 \times 5\text{-Gbps}$ e-link	13	$4 \times 5\text{-Gbps}$ e-link

In the BH, each digitized sample will be transferred to the off-detector electronics at the full 40 MHz rate, with no pipeline in the front end. As in the Phase I HE upgrade scheme [33], the light signal from each detector channel will be read out by silicon photomultipliers (SiPMs) inside RBXes located around the periphery of the highest-z layers. The channel geometry will be defined to provide longitudinal as well as  $r$ - $\phi$  segmentation. After SiPM readout, the signals will be digitized in a charge-integrating ASIC, formatted and serialized in a radiation hard FPGA,



and finally transmitted to the service cavern by 10 Gbps optical links. The charge-integrating ASIC, the successor of the current QIE11 [79] chip used in the Phase I HE upgrade, is expected to be a custom designed chip, while the FPGA serializer is expected to be a commercially available product. Buffering of data while waiting for a Level-1 accept is done in the FPGAs. Each of the 36 megatiles will be served by 32 optical links, leading to a total of approximately 1200 optical links needed in the BH, compared to 3000 in the FH (and 13,000 in the EE).

Trigger primitives, to be sent to the Level-1 calorimeter trigger, will be constructed in the off-detector electronics from the single-plane sums and the BH samples. The calorimeter electronics will be responsible for forming local longitudinal clusters and projective “towers” for use in the calorimeter trigger.

Electromagnetic clusters will be formed beginning with energetic seeds and applying a priori information about the lateral and longitudinal shape of electromagnetic showers. Layer-by-layer pileup and non-clustered-energy corrections can be made in the calorimeter electronics before transmission of the clusters to the trigger electronics. The trigger electronics will be responsible for merging clusters and carrying out isolation and E/H requirements as well as integrating tracker information where such information is available.

Preliminary studies of trigger algorithms suggest that, for the electromagnetic trigger, the use of the information from longitudinal granularity and fine lateral granularity will provide effective rate reduction. Comparing rates between the Phase 1 trigger running at an instantaneous luminosity resulting in a mean of 40 pileup interactions per bunch crossing, and the HGCal with a mean of 140 pileup interactions per crossing, the Level-1 single electromagnetic trigger rate is only 1.5 times higher for thresholds up to 20 GeV, and 2.5 times higher for thresholds up to 30 GeV, despite the increase in the instantaneous luminosity by a factor of 3.5. This is shown in Fig. 6.20. The algorithm results in only a 1–2% inefficiency for electrons.

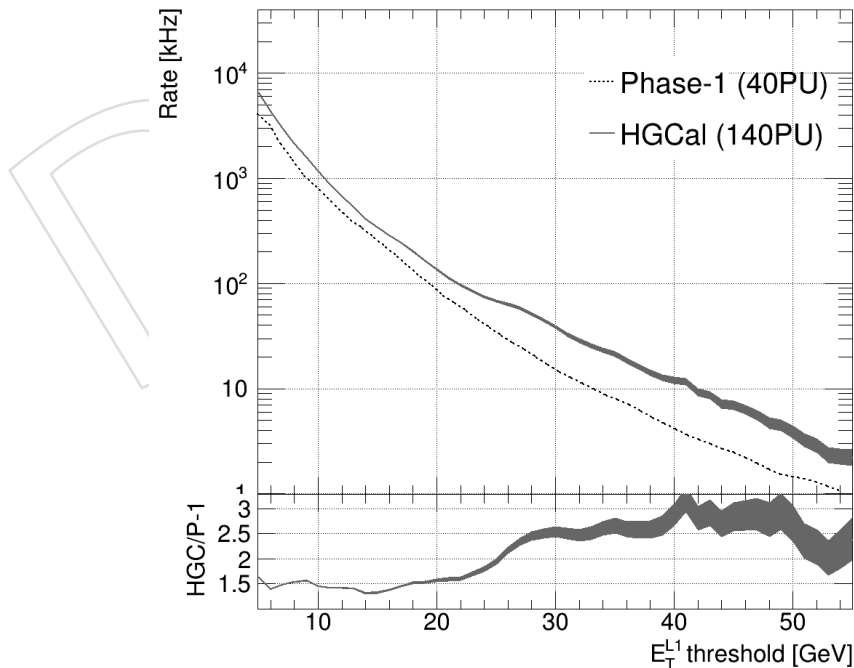


Figure 6.20: Comparison of rates between the Phase 1 electromagnetic object trigger running at an instantaneous luminosity resulting in a mean of 40 pileup interactions per bunch crossing, and the HGCal with a mean of 140 pileup interactions per crossing.

The FH and BH data will be processed in calorimeter electronics of the same design as for the EE part. Clusters from the hadronic section of the calorimeter will be sent to the calorimeter trigger electronics for use in the electromagnetic and jet algorithms. Clusters will be similar in size to the Phase-I trigger towers ( $0.087 \times 0.087 \Delta\eta \times \Delta\phi$ ) or smaller to allow for precisely-defined jets and isolation regions.

Ongoing exploration of possible jet algorithms shows that the lateral granularity allows the dense core of jets, particularly jets from the hadronic decay of  $\tau$  leptons, and quark jets from the VBF production of Higgs bosons, to be identified and distinguished from pileup. The fine lateral granularity provides effective rejection of background from pileup, where regions of high activity containing the energy from many overlapping interactions are reconstructed as jets. VBF jets and jets from the hadronic decay of tau leptons can be seeded with more than 90% efficiency using small clusters in EE and FH. An average of 16 seeds per event, providing regions of interest, are found in events where the mean number of interactions per bunch crossing is 140. Optimization of the size of the  $\eta$ - $\phi$  region used for reconstructing jet energy has been studied.

A size of  $\Delta R = 0.2$  seems to give the best balance between reduction of pileup fluctuations and the minimization of fluctuations due to inadequate jet containment. Cells are included in the energy sum based on a layer-by-layer pileup-dependent threshold, and the resulting jet energy is corrected for pileup and out-of-cone leakage. Figure 6.21 compares the Level-1 single jet trigger rate in the endcaps for the Phase 1 detector at an instantaneous luminosity resulting in a mean of 40 pileup interactions per bunch crossing, and the HGCal, using the algorithm described, with a mean of 140 pileup interactions per crossing. The rate in the HGCal is at most a little more than twice the rate in the Phase 1 endcaps, despite the instantaneous luminosity being 3.5 times higher.

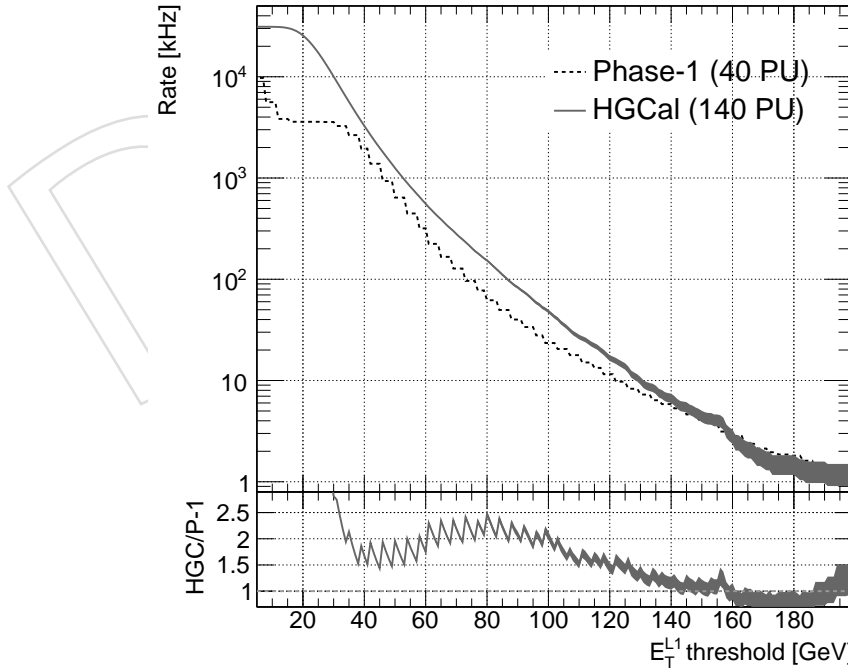


Figure 6.21: Level-1 single jet trigger rates in the endcaps, comparing the Phase 1 trigger running at an instantaneous luminosity resulting in a mean of 40 pileup interactions per bunch crossing, and the HGCal with a mean of 140 pileup interactions per crossing.

Besides the trigger electronics, several other off-detector electronics systems will be required. While trigger data and the full BH data will be available off-detector for every beam crossing, the full EE and FH data will be transferred upon receipt of an L1A to DAQ boards which will interface between the calorimeter electronics and the central DAQ system. Robust control and safety systems will also be required to configure, control, and protect the detector.

The estimated power consumption of the FH detector elements (excluding off-detector electronics located in the service cavern) is 30 kW, with  $\sim 20$  kW consumed by the front end readout ASICs. At the 1.2 V level required by the ASICs, this amounts to approximately 25 kA of current. Heavy duty copper tracks on a 0.5 mm thick PCB are used to take the low voltage power from the cassette periphery to individual modules within the cassette, as shown in Fig. 6.11. A section of about 50 mm<sup>2</sup> of copper per cassette is needed.

DC-DC converters are used to reduce the required cross-section of the cables between the power supplies and the detector. These are located outside the cold volume, in the low radiation region at the back of the endcap calorimeter structure on its outer circumference ("RBX region"). The cross-section of these cables, which run a length of 3-4 m, has been dimensioned to limit the voltage drop to between 10% and 15% (total), and voltage regulators on the detector modules are foreseen and accounted for in the power estimates.

The cable runs from the power supply racks on the experimental cavern balconies to the DC-DC converters at the RBX region have a typical length of 10 m, and the longest path from the DC-DC converters to the HGC modules is about 4 m. Based on the ongoing development of DC-DC converters for the CMS Phase I pixel upgrade [34] and Phase II outer tracker upgrade, we assume a conversion ratio of 8/1 with a 65% efficiency. Under these assumptions, and allowing for a power loss over the cables consistent with an overall 50% power delivery efficiency (i.e. including loss in DC-DC converters), the copper cross-section required for the entire HGCAL from the power supplies to the DC-DC converters and through the services choke points across the ME1/1 chambers is about 60 cm<sup>2</sup>, with  $\sim 20$  cm<sup>2</sup> devoted to FH (**assuming the number of cables from the balconies to the RBX region scales as the number of layers in HGCAL**). A copper cross-section of about 75 cm<sup>2</sup> is required (in FH only) for the cables from the DC-DC converters to the front-end modules (about 1-2 mm<sup>2</sup> per module). The total cross section required for the entire HGCAL is 200 cm<sup>2</sup>.

The BH RBXes will be powered from supplies located inside the experimental cavern. DC-DC converters located inside the RBXes will regulate the supply voltage to the levels required by the BH scintillator readout electronics (between 1.2 and 5 V). **These values come from the Phase I HCAL upgrade TDR. Each of the 36 BH RBXes is expected to dissipate 300 W of power, leading to a total power consumption of 10 kW for the BH portion of the HGCAL. Is this true for Phase II? It comes from the Phase I TDR. Assuming the channel count remains the same for Phase II but all the o-links are upgraded from 5 Gbps (400 mW/link) to 10 Gbps (750 mW/link), the power dissipation per RBX would only increase by 4%.**

### 6.3.7 Calibration and monitoring

The sensor intercalibration will be tracked and maintained using the MIP signals in any triggered event, to follow the slow change in charge collection efficiency over the duration of HL-LHC operation. The noise performance of the ToT front-end will enable MIP signals to be seen and fitted in almost every sensor cell of the HGCAL even after exposure to the hadron fluence from an integrated luminosity of 3000 fb<sup>-1</sup>. In addition, for redundancy, and to fully guarantee the ability to achieve MIP calibration throughout the life of the HGCAL, dedicated low-capacitance/low-noise cells will be included on each wafer. For these calibration cells a

standard hexagonal cell will be divided into seven subcells, a central hexagonal subcell with six neighbours thus providing cells with a signal-to-noise ratio  $> 5$  on each wafer of the HGCAL, even after the accumulation of the full lifetime hadron fluence.

The electronics chain of each channel will be independently monitored and linearized using a charge injection system based on a chopper circuit and fixed calibration capacitances connected to the front-end input. The additional noise contribution from inclusion of this system is negligible because of the chopper series resistance. The large range of injectable charges (0–10 pC) will allow a finely detailed linearity measurement. Before installation, the channel testing and qualification protocol will include measurement of the charge injection capacitance to  $< 1\%$ . Knowledge of the charge injection capacitance will allow an independent verification of the MIP calibration at startup, and, when used in conjunction with the MIP calibration, will monitor the charge collection efficiency throughout the lifetime of the detector.

As in the current hadronic calorimeter, BH WLS fiber response will be monitored frequently by LED light injection during run periods. During long shutdowns, radioactive sources can be moved into the vicinity of the scintillator tiles to monitor the full scintillator + fiber response. Loss of light yield in the scintillator and WLS fibers induced by hadron damage will be calibrated out using measurements from the LED and sourcing systems, as is done for the current HCAL.

## 6.4 US deliverables

### 6.4.1 Silicon sensors and modules

### 6.4.2 Scintillator modules

### 6.4.3 Hadron calorimeter cassettes

### 6.4.4 Electronics and electrical systems

## 6.5 Performance

- 7 pp.
- Occupancy
- Effect of dead time due to large signals
- Determination of shower direction (unique feature of HGCAL w.r.t. old EE/HE)
- Jet energy resolution
- Rejection of PU jets
- (Maybe) Muon ID

The FH hadronic calorimeter described in Section 6.3 has been simulated, along with its EE electromagnetic counterpart, in HL-LHC scenarios corresponding to 140 and 200 pileup interactions per bunch crossing. These simulations indicate that the relative energy resolution (Section 6.5.1) for EM and hadronic objects will be sufficient to meet the demands of HL-LHC physics analyses. In addition to its performance as a calorimeter, the simulations also demonstrate the operability of the proposed HGCAL detector in HL-LHC conditions. It is expected that the primary design goal of sensitivity to single MIPs (Section 6.5.2) after receiving a dose consistent with  $3000 \text{ fb}^{-1}$  of integrated luminosity can be met. Furthermore, the detector read-out scheme is capable of handling the expected channel occupancy (Section 6.5.3).

Prototype HGICAL slices were operated in test beams at Fermilab ( $0.6X_0$ - $15.3X_0$ ) and CERN ( $5X_0$ - $27X_0$ ) in 2016 (Section 6.5.4). Results show good linearity over a range of energies. The measured energy resolution approaches the design goal and is consistent with expectations for the range of  $X_0$  covered, absorber layer thicknesses, and silicon layer spacing, all of which differ from the ultimate proposed design in these initial runs. Overall, the test beam exercise testifies to the soundness of the HGICAL detection principle.

Early prototypes of the front end readout chip show good linearity over the full ToT dynamic range and low noise in the shaper circuit.

### 6.5.1 Design energy resolution

### 6.5.2 MIP sensitivity

Requiring a signal  $> 0.9$  MIP in the layers before and after the sensor under study enables a clear peak to be seen and fitted for noise levels up to about 0.4 MIP. Figure 6.22 shows the result in sensor cells with noise equivalent to 0.3 MIP located in the region  $2.8 < |\eta| < 2.9$ . Local isolation, requiring that all sensor cells surrounding the cell under study have a signal  $< 0.5$  MIP, reduces the sensitivity of the fitted peak value to the instantaneous luminosity. The sensors in this region have an active thickness of  $100\ \mu\text{m}$  and the simulation is made with a mean of 140 interactions per bunch crossing.

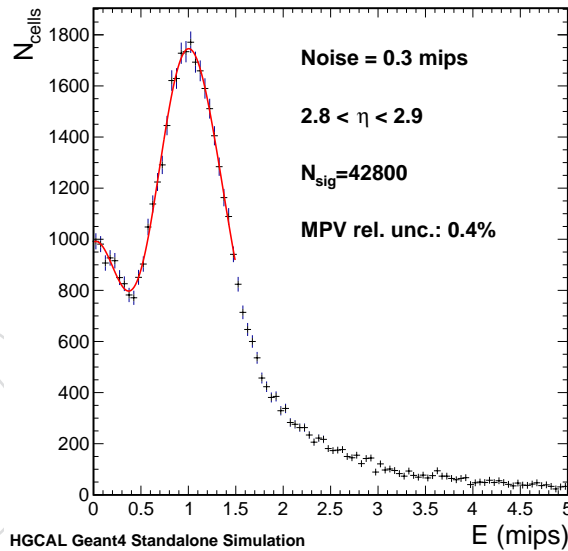


Figure 6.22: One-MIP signal peak in sensor cells with noise equivalent to 0.3 MIP, after MIP tracking and local isolation algorithms have been applied. The sensor cells are located in the region  $2.8 < |\eta| < 2.9$  and have an active thickness of  $100\ \mu\text{m}$ . In the simulated events the mean number of interactions per bunch crossing is 140.

Tightening the MIP-tracking algorithm by requiring corresponding signals in two layers before and two after the sensor under study removes the contribution from instances where there is, in fact, no MIP present (seen on the left of the signal peak in Fig. 6.22) and enables the peak to be seen and fitted even when the noise level is 0.6 MIP. This noise level corresponds to that in the worst case sensor pads after a fluence corresponding to  $3000\ \text{fb}^{-1}$ .



### 6.5.3 Occupancy

The occupancy has been studied in events where the number of pileup interactions is modeled by a Poisson distribution with a mean of 200. The result is shown in Fig. 6.23 for signals above a threshold of 0.5 MIP, equivalent to an energy of 2.5 MeV, and for signals above a threshold of 5 MIPs. The occupancy for 0.5 MIP threshold is used to estimate the data rates per module at 1 MHz L1A rate shown in Table 6.2, which are well within the proposed optical link bandwidth. Occupancies for 140 pileup interactions per crossing can be found by scaling these occupancies by a factor 0.7 ( $= 140/200$ ).

The ToT front-end results in dead time for sensor cells with signals larger than than 250 fC. The length of the dead time depends on the signal size, and extends to 270 ns ( $\sim 10$  bunch crossings) for a cell with 210 pC deposited. The dead time has been investigated using a simulation of a sequence of ten bunch crossings for instantaneous luminosities corresponding to means of 140 and 200 interactions per bunch crossing. The resulting probability for a sensor pad to be busy during any particular bunch crossing, as function of layer, for different  $\eta$  locations, is shown in Fig.6.24. It is found that the probability for a sensor cell to be busy is well below  $\times 10^{-4}$  everywhere.

### 6.5.4 Test beam results

## 6.6 Alternatives

- 3 pp.
- Alternatives under consideration within BH: SiPM-on-tile
- Alternatives previously considered: liquid scintillator, etc
- Replace highest  $\eta$  ring with Si active medium
- 6-in. silicon (conservative option) jettisoned for 8-in. (cost savings, technology now demonstrated by Hamamatsu)
- n-in-p everywhere vs. p-in-n at lower fluence areas (cost savings)
- Electrical links from cassettes vs. optical links

## 6.7 R&D Towards Final Design

- 3 pp.
- Low priority topic initially due to project evolution

### 6.7.1 Radiation hard plastic scintillator

Efforts to identify a plastic scintillator for the BH megatiles meeting HL-LHC requirements is proceeding on three fronts: irradiation of candidate tile samples *in situ* on the CASTOR table of CMS during Run II, dedicated high dose gamma and neutron irradiation campaigns, and research into doping techniques during production. Scintillator light output decreases exponentially with integrated dose according to the formula

$$L(d) = L(0) \exp d/D \quad (6.1)$$

where  $L$  is the light output after integrated dose  $d$  and  $D$  is the dose constant. Dose constants have been measured for a variety of types of plastic scintillator at a variety of different dose rates. Data at low dose rates generally come from long running particle physics experiments in

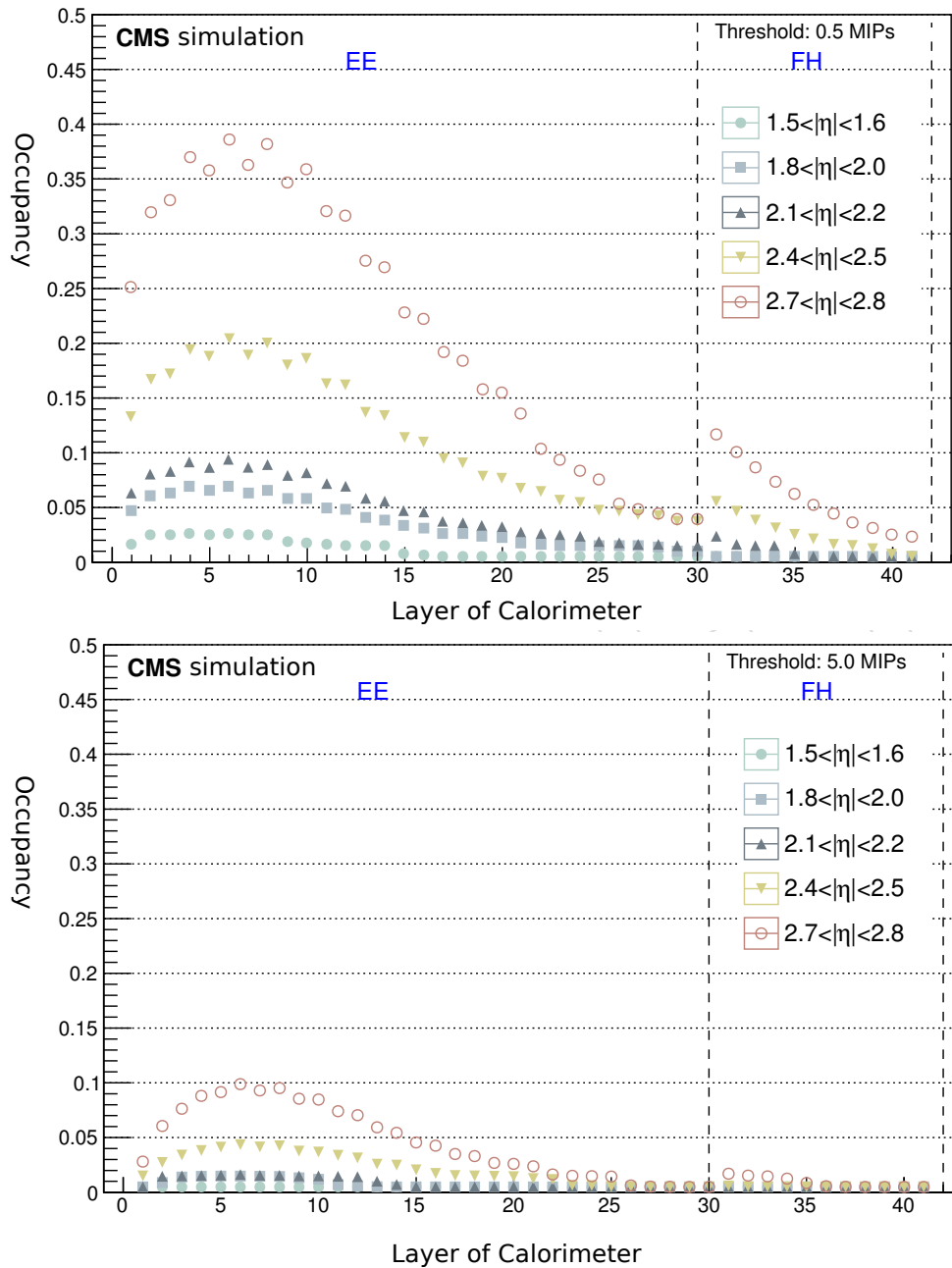


Figure 6.23: Median cell occupancy as a function of layer for signals (top plot) above 0.5 MIP and (bottom plot) above 5 MIP for an instantaneous luminosity such that the mean number of pileup interactions per bunch crossing is 200.

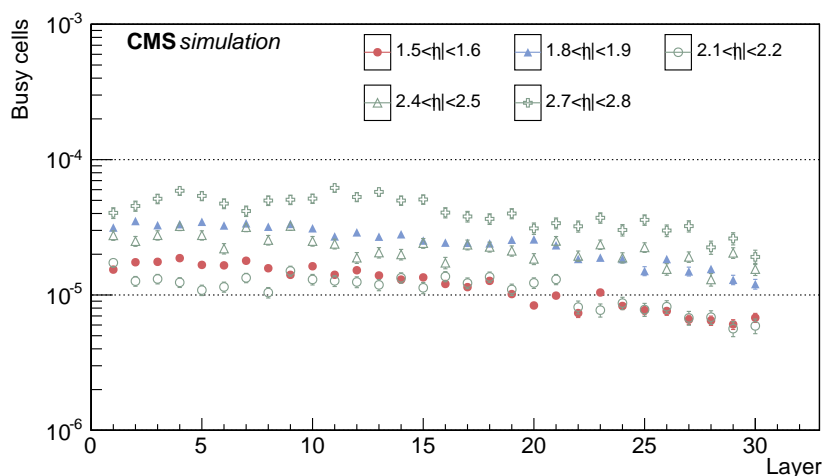


Figure 6.24: Probability for a sensor pad to be busy during any particular bunch crossing, as function of layer, for different  $\eta$  locations, when the instantaneous luminosity is such as to result in a mean of 140 pileup interactions per bunch crossing.

hadron beams, while data at high dose rates generally come from dedicated sample irradiation campaigns. Dose rate is a critical parameter for plastic scintillator, as the relative rate of oxygen diffusion (controlled by production techniques) and radical creation (controlled by dose rate) changes how the radiation damage occurs. Presence of oxygen tends to result in enhanced damage of the substrate and/or the primary fluor, resulting in higher damage for a lower dose rate, given the same total dose. Secondary fluors may exhibit a reverse dependence on oxygen. Figure 6.25 summarizes neutron and photon irradiation measurements on a variety of plastic scintillators at different dose rates.

## 6.7.2 Trigger data compression

## 6.7.3 Cassette prototyping program

## 6.8 Risk

Editor(s): J. Mans

- 2 pp.

## 6.9 Value Engineering

- 2 pp.

## 6.10 Quality Assurance and Quality Control

- 3 pp.

### 6.10.1 Sensor and Module Quality Assurance and Quality Control

A common procurement process for the HGC sensors across the international CMS Phase 2 upgrade project is expected. The US will be responsible for purchasing the 8500 (??) sensors

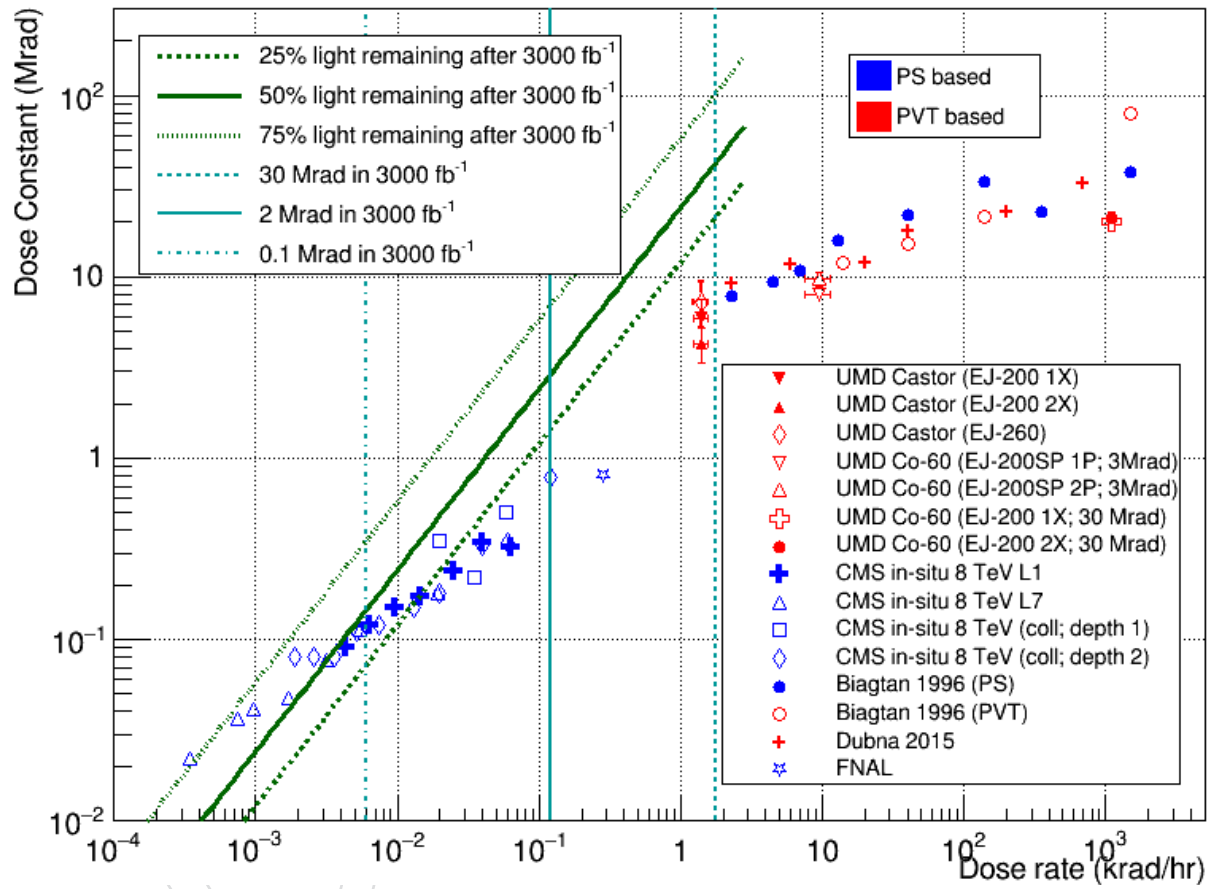


Figure 6.25: Markers: Dose constants of different types of plastic scintillator measured at different dose rates. Dark green lines: necessary dependence of dose constant on dose rate to achieve a given light loss after  $3000 \text{ fb}^{-1}$  HL-LHC integrated luminosity. Any scintillator “above” the green lines is sufficient for BH. Teal lines: average dose rates corresponding to a given dose integrated over  $3000 \text{ fb}^{-1}$  of HL-LHC operation.

for FH corresponding to 20% (??) of all sensors to be acquired for the HGC. The vendor will be responsible for the sensors to meet specifications and only 5% of additional sensors are assumed as spares for damage or loss during shipment, module assembly, and other processing steps. The vendor risk is controlled by a general requirement of limiting the fraction of the vendor's capacity and sales volume used by the CMS orders to about 10%, where the entire HGC order may also be split between multiple vendors.

Quality assurance will be performed on batch samples of sensors. Standard QA procedures to be followed for all sensors will be developed with an emphasis on the R&D to obtain a sensor design that maximizes the production quality. The use of a pre-series production will help to find additional problems for QC. During the R&D phase a thorough testing of sensors including IV- and CV-curves for all channels plus interpad capacitance measurements are foreseen. For the sensor production phase the testing of several categories of pads (inner pads vs. outer pads) is assumed to be sufficient. All sensor test results are recorded into a database.

Regarding module production, experiences with the CMS Outer Tracker construction for Run 1 and with the EC prototypes for the beam tests at Fermilab and CERN inform our QA/QC procedures. Common-mode noise, poor quality wire-bonds, broken PCB traces and wires, non-uniform epoxy layers between the layers, epoxy filled holes in PCBs, degradation of (conductive) silver epoxy, misalignment of layers and similar issues have emerged and proper actions have been taken in each case. An updated and detailed list of issues and their solutions are being documented throughout the initial R&D process in order to establish quality assurance.

During the module assembly process, the faults need to be identified quickly for feedback to the production line. Each step in the assembly line will be documented and updated as needed. Quality control will be established after each major step (e.g. dimensional evaluation after the components are glued into a stack, periodic wire-bond strength tests after the PCBs are wire-bonded to sensors, and a suite of electrical/readout tests of every module after the cover plate is installed). We will establish a database and use results of many partially correlated tests to determine the type and location of faults. The assembled modules will be bar-coded, packaged and temporarily stored before shipment in custom containers for cassette assembly.

## 6.11 Environmental Impact, Health, and Safety

Editor(s): R. Yohay

- 0.5 pp.
- Cryogenics
- Radioprotection
- CO<sub>2</sub>—ODH?
- Heavy machinery (cassettes are 200 kg)



DRAFT

## Chapter 7

# Trigger and DAQ - Editor Berryhill 40 pages

## 7.1 Introduction

- Berryhill, 1.5pp

*List of items from Jeff: a description of the basic problems and strategies for trigger and data acquisition in HL-LHC conditions; . basic external constraints; basic physics requirements; Run 2-preserving menu; why we need improved algos and tracking in addition to more bandwidth and latency; brief descriptions of technology solution: fpgas on boards in crates connected by fibers; enumeration of sections*

The current CMS trigger system is designed to reduce the rate of selected events for physics analyses from approximately 40 MHz produced in pp interactions at the LHC down to 400-600 Hz to be stored for physics analysis and consists out of the first level (L1) and the high level (HLT) triggers. The L1 trigger is implemented in hardware, and selects events based on detector signals consistent with electrons/photons, muons,  $\tau$ -leptons, jets, or based on some global quantities. The trigger thresholds are adjusted to restrict the L1 output rate to 100 kHz, the upper limit imposed by the CMS readout electronics. The HLT is implemented in software should further reject events resulting in an average rate of 400 Hz (with maximum up to 1 KHz) for offline event storage.

The current L1 trigger uses information from calorimeter and muon system and has a fixed latency, 4  $\mu$ s, during which the system must decide if an event should be tentatively accepted or rejected. The angular coverage of the calorimeter trigger is restricted to  $|\eta| < 5$ , while the muon system covers region  $|\eta| < 2.4$  **FIXME:** ?? is this correct.

The L1 trigger uses high-speed custom application-specific integrated circuits (ASICs). **FIXME:** May be we should write something here about speed etc ??

During the HL-LHC operation the maximum number of pp interactions within one bunch crossing (PU) can reach 200, compared to approximately 30 in the current LHC operation. It should result in increase of the L1 rate up to 4000 kHz, significantly beyond the technical feasibility of the CMS trigger system. The most effective way to reduce the PU contribution is to use information from tracking systems that allows not only properly reconstruct the primary vertex, but also to differentiate between tracks coming from the primary vertex and PU events, thus reducing the rate of events by factor of 10, compare to L1 without tracking system included.

With including tracking information the L1 latency increases up to 12.5 sec to provide sufficient time for hardware to reconstruct tracks and to perform matching with the calorimeter and muon system. With the L1 acceptance rate of about 500-750 kHz and assuming that about a 1/100 event selection is maintained from L1 to HLT output, a permanent event storage of 5-7.5

4177 kHz rate will be required to operate at 200 PU.

4178 Besides the addition of a Tracker Trigger at L1, which is an integral part of the design of the new  
4179 silicon tracker, CMS also plans an upgrade to the detector readout, L1 trigger and HLT systems,  
4180 which would allow up to 750 kHz L1, up to 12.5 sec L latency, up to 7.5 kHz permanent event  
4181 storage rate.

4182 **FIXME:** May be we want to say here few words about calorimeter segmentation, single crystals  
4183 and also about extended eta coverage that is essential for physics ?? Should we say few words  
4184 about FPGA here??

4185 The main contributions of the US Universities to the L1 trigger upgrade are: **FIXME:** Here we  
4186 need to put US deliverables that should correspond BOE !!! i.e. explain exactly what we are  
4187 doing and will do. I wrote something what I think is essential for calorimeter trigger

- 4188 • Track trigger
- 4189 • Calorimeter trigger: development of the calorimeter L1 trigger architecture based  
4190 on the simulation studies, development of all components of hardware, firmware  
4191 and software, including testing and operational support of the calorimeter trigger  
4192 system in situ.
- 4193 • Muon trigger
- 4194 • Correlator This part of the system should collect the information for the systems to  
4195 provide final information for the HLT.

## 4196 7.2 Requirements

4197 **Editors: Development to be led by Chris Hill, Berryhill, Cavanaugh, 5pp**

4198  
4199 list of physics requirements for the trigger.

4200 emphasize rate/threshold/efficiency goals for each single object (or essential cross triggers,  
4201 such as diphoton or muon-tau)

- 4202 • muon

4203 The overall goal of the lepton triggers, and the muon trigger specifically, is to main-  
4204 tain sensitivity for electroweak scale physics and for possible new physics despite  
4205 the higher luminosity and pile-up conditions of the HL-LHC. This translates into  
4206 single lepton thresholds below about 30 GeV, with a rate acceptable for the over-  
4207 all physics menu, in order have good acceptance to final states with a leptonic W  
4208 decay, e.g. in Higgs or top decays. The measurement of muon  $p_T$  using only the  
4209 muon detectors has a limited resolution (of order 20%) and thus leads to a trigger  
4210 rate that flattens at higher thresholds. In order to improve on the resolution, and  
4211 thus the steepness of the rate curve versus threshold, a matching to “tracker tracks”  
4212 found in the inner silicon system is required such that the resolution is improved  
4213 to 1–3%. Therefore, the first requirement of the muon trigger upgrade is to identify  
4214 and report muon tracks reconstructed standalone in the muon detector systems in  
4215 the convention necessary to facilitate the correlation with tracker tracks.

4216 A second requirement is to maintain high efficiency for identifying muons. Aside  
4217 from the obvious advantage of maximizing the utility of the delivered high lumi-  
4218 nosity of the HL-LHC for physics, the inefficiency for multi-lepton triggers is com-

pounded by the inefficiency per lepton leg. Thus the goal is to main a single muon efficiency above 80%. Since the forward muon system currently has only one detector technology for  $|\eta| > 1.6$ , cathode strip chambers (CSCs), the HL-LHC upgrade of CMS envisions the addition of new detectors in this region to improve upon the redundancy for an efficient trigger, and to offer improved standalone momentum measurements [cite TP][? ]. Specifically, gaseous electron multiplier (GEM) detectors are planned for the first two disks of the endcaps (GE1/1 and GE2/1) as well as for a high- $\eta$  tagger (ME0), and resistive plate chambers (RPCs) for the third and fourth disks, as shown in Fig. 7.1. Thus the efficiency requirement translates into a requirement on the information processing of the muon trigger, which must be expanded to accept the new detector signals and to use them in the track-finding logic.

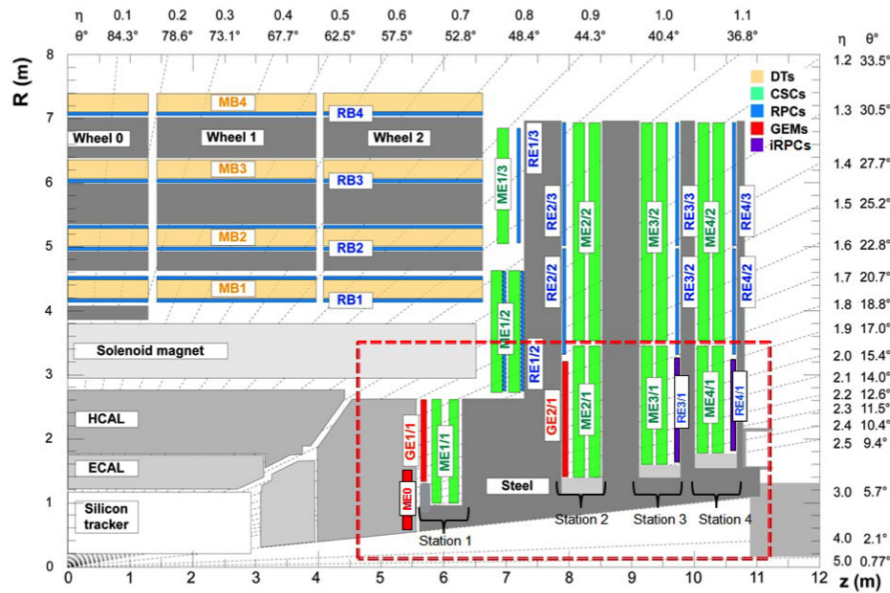


Figure 7.1: Quarter cross-sectional view of the CMS muon detector systems with the new HL-LHC detectors highlighted.

A third physics requirement is to open the acceptance to physics signals beyond the standard model, particularly to muons originating from a new long-lived particle that leads to a significantly displaced vertex ( $>$  several mm). While the current muon trigger can have acceptance to such signals, it's momentum assignment is severely biased since it assumes a vertex constraint. This leads to high rate for any  $p_T$  threshold. Thus the upgraded muon trigger must have track-finding (or pattern) logic to identify displaced tracks, and track-fitting logic to assign the momentum without a beam constraint.

- electron
- tau
- photon
- jet
- missing energy
- HT

Sascha : I put here something in a little bit different format, I would start the section just saying what

*we want to achieve, put requirements on the objects and on the general performance and after that what it requires from the subsystems*

The purpose of the calorimeter trigger system is to identify electron, photon, jet and  $\tau$ -candidates, based on information from the ECAL and HCAL detectors across the entire range of CMS coverage in  $\eta$ . In addition, the calorimeter trigger computes global energy sums (Total  $E_T$ ,  $E_{miss}$ ,  $H_T$ ), and is capable of providing a real-time estimate of bunch-by-bunch background levels as an input to pileup-dependent trigger selections.

Precision measurements of Higgs properties and other electroweak phenomena are central goals of the HL-LHC physics programme and can only be attained by efficiently accepting events that have photons or leptons with low transverse momentum of order of 20 GeV and loose identification requirements. To provide improved performance of the trigger better position resolution of the calorimeter objects for matching with tracks is required, as well as sharper turn-on efficiency curves with efficiency of 97-99% at the plateau. Such improvements are critical, for example, in efficiently selecting Higgs decays to offshell Z bosons, which can have multiple leptons in the final state that often have low transverse momenta and/or are not well isolated. Searches for new physics often also involve low pT leptons in the cascade decays of possible heavy new particles.

The summary of the key requirements for the Phase-II Calorimeter Trigger system:

- spatial resolution of electromagnetic clusters reconstruction as close as possible to the offline reconstruction,  $0.015 \times 0.015$  in the  $\eta \times \phi$  phase space ;
- ability to reconstruct electromagnetic clusters with  $P_T$  of few GeV with efficiency above 95% in the region above 10 GeV (? need to think about a number !!!) ;
- should have latency for trigger processing about (not more then)  $4 \mu s$  ;
- the format of the calorimeter trigger output should allow combination with the tracking information ;
- provide detailed information for use in muon isolation calculations ;
- should provide identification of jets, either via multiple fixed-sized window algorithms, or an iterative clustering algorithm

## 7.3 Proposed design

- 15pp

Go over the technical design for each subsystem. How many stages, how many boards with what I/O, memory/ASIC, and FPGA requirements. What algorithms are executed where.

### 7.3.1 Architecture Overview

- Berryhill, 1.5pp

### 7.3.2 Calorimeter trigger

- Savin, 3pp

The present ECAL electronics has maximum latency of  $6.4 \mu sec$  and approximately 150 kHz of level 1 acceptance rate. The concept for new electronics is to read out single crystal information



at 40 MHz, moving all trigger primitive generation and buffers off-detector. This will allow the system to operate without on detector L1-acceptance rate limitation and the off-detector data buffers will provide sufficient latency for the new trigger.

The HCAL Phase-I upgrade foresees that by the end of LS2 the entire HCAL back-end electronics will be  $\mu$ TCA-based. For HL-LHC it is foreseen that HCAL performs continuous (40 MHz) readout from the detector front-end to the back-end, where the data is inserted in a pipeline awaiting for the trigger decision. The foreseen rate of 750 kHz could be achieved by a redesign and replacement of the back-end cards. However, since EE and HE will be replaced for Phase-II and the front- and back-end of EB will be redone, a more-sensible scenario foresees that the entire new endcap calorimeter and both EB and HB would have newly designed back-end electronics systems. In this scenario, the existing Phase-I  $\mu$ TCA HCAL back-end electronics would be fully devoted to HO and HF. With the increased availability of  $\mu$ TCA components for HF and HO, achieving a 750 kHz rate will become possible.

Following the upgrade of the on-detector and off-detector electronics of the barrel ECAL and HCAL, substantially enhanced input data will be available for calorimeter trigger. The digitized response of every crystal of the barrel ECAL (instead of the present  $5 \times 5$  crystal sums) will provide energy measurement at a spatial resolution of  $0.0175 \times 0.0175$  in  $(\eta, \phi)$ , as opposed to the current input to the trigger consisting of trigger towers with granularity  $0.0875 \times 0.0875$ .

The purpose of the calorimeter trigger system is to identify electron, photon, jet and  $\tau$ -candidates, based on information from the ECAL and HCAL detectors across the entire range of CMS coverage in  $\eta$ . In addition, the calorimeter trigger computes global energy sums (Total ET, Emiss, HT), and is capable of providing a real-time estimate of bunch-by-bunch background levels as an input to pileup-dependent trigger selections. Important goal of the Phase-II L1 trigger is to produce objects as close as possible to those used in the online selection and in the physics analyses. With the Phase-II upgrade, candidate particle objects, such as electrons, photons, and taus, with position, energy, and isolation characteristics close to those of the final particles will provide sharper turn-on efficiency curves.

An increased latency of up to  $4 \mu\text{s}$  will be available for stand-alone calorimeter trigger processing, compared to less than  $2 \mu\text{s}$  in the pre-LS3 system. This will allow the implementation of object-finding algorithms with far greater selectivity than today, making full use of the improved input data.

The calorimeter trigger will be incrementally upgraded in the period between LS1 and LS2, with new  $\mu$ TCA-based hardware providing significantly higher processing capacity than in the pre-LS1 era. However, it is envisaged that the HL-LHC trigger will be based on entirely new state-of-the-art hardware, and make use of an architecture that brings together calorimeter, muon and track trigger information at the earliest possible stage. Since the implementation will make use of flexible generic modules, there are several possible options for the mapping of calorimeter trigger processing steps onto hardware.

The proposed structure of the calorimeter trigger system is shown in Fig. 7.2. The data are processed as soon as they come, crossing-by-crossing data flow supports multi-pass algorithms as pipeline operations, the data can be kept in the FPGA as long as necessary. It places less restrictions on calorimeter data volume.

The architecture requires 2 layers. The cards for both layers are identical and allow for multiple pass on the same data, for calibration, correction or combination purposes.

For current cost and number of cards estimate the existing CTP7 card is used. The card has

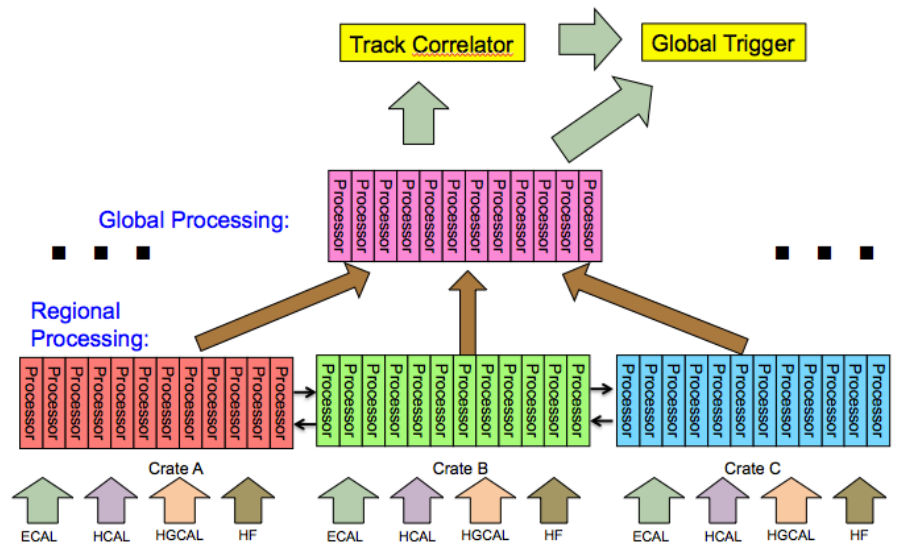


Figure 7.2: Proposed structure of the calorimeter trigger system.

Virtex-7 690T FPGA data processor together with ZYNQ 045 SoC device, 12 MGT  $\mu$ TCA back-plane and serves 67 Rx and 48 Tx 10G optical links. In the future the FPGA capabilities are expected to change with possible increase of the cards costs. Number of input channels and needed cards for parallel design of the calorimeter trigger system is summarized in Table 7.1.

A new module - the Advanced Processor module is under design now, it is an FPGA Processing Platform targeted for HL-LHC trigger and back-end electronics. Features include 100 Optical I/O connections, multi-rate capable Channel count driven by FPGA packaging, 1-10+ Gbps link base range with 25G extended range, capability for an expansion mezzanine for memory LookUp Tables (LUTs) and/or a supplemental FPGA processor, an embedded Linux control platform, customizable I/O capability, and scalable costs with different FPGAs. The prototype of this module should be build together with demonstrator, pre-production prototype and finally production following this design.

Table 7.1: Number of input channels and boards for parallel calorimeter trigger processing

ECAL Barell channels	61200	for now number of crystals
ECAL Endcap channels	61000	may change for HGAL case
HCAL Barell/Endcap	13824	the same as in Phase 1
HF	1728	as above but combining 2 PMT
Information per channel (bit)	12	10 energy + 2 quality bits
Total bits	1653024	
Bandwidth (bits/s)	$6.61 \cdot 10^{13}$	data transmitted at 40 MHz
Bits/s card	$4.92 \cdot 10^{11}$	assuming 80x10 Gbps links
		running 192 bits at 40 MHz
		with 80% packing efficiency
Number of cards Layer 1	135	
Number of card Layer 2	45	
Total cards	180	

### 7.3.3 Muon trigger

- Acosta, 3pp

The architecture of the endcap muon trigger is similar to the present phase-1 upgrade currently in place, the Endcap Muon Track-Finder (EMTF). Two processing layers are envisioned as shown in Fig.7.3, one layer for standalone muon track-finding, divided regionally into azimuthal sectors with each processing element referred to as a Sector Processor, and a second Muon Global Sorting layer for sorting, duplicate cancellation, and global parameter assignment. The number of processing nodes in both layers is increased by a factor two from the present Endcap Muon Track-Finder (EMTF) to accommodate the increased input bandwidth from the additional forward muon detectors and to provide additional processing power for the dense HL LHC environment. This breaks down to 24 processing nodes for the first processing layer (12 per endcap), and 2 for the second layer (1 per endcap).

The details of the muon detector data input bandwidth are shown in Tab. 7.2. No further data concentration of the legacy CSC and RPC optical links is assumed. The total sum is 1164 detector links with link speeds varying from 1.6 Gbps to 6.4 Gbps, for a total input data bandwidth of about 5 Tbps. Approximately 25% of links will need to be passively split to ensure efficient data coverage across processor boundaries according to the current EMTF design. Assuming also an 80% data packing fraction into the link frames leads to a total of about 1800 total fibers connecting to the Sector Processors of the track-finding layer. Assuming that each Sector Processor can accept up to 80 fibers as the current EMTF leads to the need of 24 Sector Processors in the first processing layer.

Track-Finding Layer

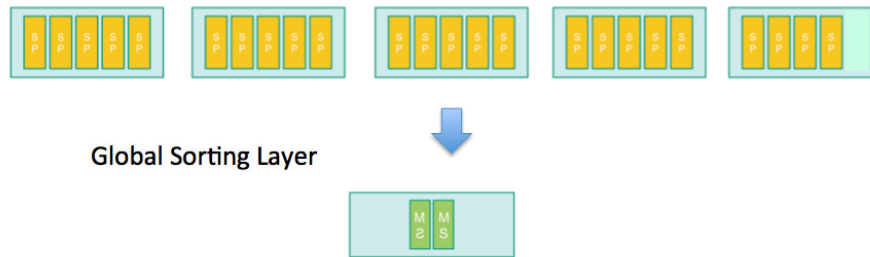


Figure 7.3: Proposed structure of the muon trigger system.

Table 7.2: Optical link inputs from the muon detectors to the Muon Track-Finder.

Detector	Link speed (Gbps)	total links
CSC	3.2	480
RPC ( $ \eta  < 1.6$ )	1.6	180
GEM ME0	6.4	72
GEM GE1/1	6.4	144
GEM GE2/1	6.4	144
RPC 3/1	6.4	72
RPC 4/1	6.4	72

We conservatively assume 5 Sector Processors per ATCA crate for the first processing layer (5 ATCA crates), and a sixth ATCA crate for the Muon Global Sorting cards.

Each processing card contains a large FPGA with a large number of high-bandwidth serial link inputs. The first layer Sector Processors implement coordinate conversion in the FPGA after receiving aligned data from optical links, then perform pattern recognition to identify muons in the forward muon detector system. Patterns for both prompt muons and displaced muons

are included. Preprocessing for the momentum assignment also is performed in the Sector Processor FPGAs. Unique to the Sector Processor boards is fast access to very large memory resources ( $\sim 64$  GB) near the FPGA for use as a look-up table for momentum assignment. The processors for the Muon Sorter implement sorting logic into the FPGA.

### 7.3.4 Track trigger

- Wittich, 3pp

In order to provide tracking information to the L1 trigger, the outer tracker will utilise a novel type of module that transmits hits consistent with charged particles above a transverse momentum ( $p_T$ ) threshold of about 2 GeV, at the full 40 MHz beam crossing rate. Since about 99% of all tracks in CMS are below  $\approx 2\text{GeV}/c$ , and are of no relevance to most practical triggers, this selection reduces the volume of data the trigger will receive down to approximately 20 Tbps with an acceptable loss of low-momentum tracking efficiency. The basic concept is to compare hit strips on two closely spaced sensors of a two-layer module to reject patterns that are consistent with a low transverse-momentum track. Hit combinations in the two sensors consistent with a high- $p_T$  track segment are known as stubs and are used to form trigger tracks. In order for the tracks to be used downstream, they must be available to the correlator trigger within  $5\ \mu\text{s}$  of the collision.

In the past, hadron collider experiments have successfully made use of information from silicon-based trackers in L2 triggers based on Associative Memories (AM). However, in those applications the event rate was much lower and the latency much longer than foreseen in the CMS case. Similarly, FPGA-based applications have been used in L1 trigger applications, but at much lower channel count and precision. Thus, the challenge of developing a L1 Track Finding system for CMS is still to be met. The project will benefit from the rapid technological progress expected in the areas of FPGA processing power and data link bandwidth in coming years. Keeping up with the high stub rate will require the implementation of both time and regional multiplexing to enable parallel processing of data from different bunch crossings and detector regions.

The Associative Memory (AM) + FPGA approach makes use of a massively parallel architecture to quickly tackle the intrinsically complex combinatorics of track finding algorithms, avoiding the typical power law dependence of execution time on occupancy. The time required to perform pattern recognition is approximately linearly proportional to the number of hits. In this approach, the Tracker is divided into 48 angular regions (regional multiplexing) called trigger towers (6 in  $\eta$  by 8 in  $\phi$ ). Approximately 300 stubs per bunch crossing are expected in each trigger tower at 140 pileup. The data is formatted with coarser resolution with the detector pitch and the pattern recognition is performed by matching compatible sequences of low-resolution stubs in each detector layer with to be matched against about one million reference track patterns stored in custom-designed AM chips. An estimated 100 million reference patterns are needed for the full Tracker. The matched patterns are then retrieved and the associated high-resolution data is then subject to a second stage of pattern recognition with a much less severe combinatorial problem. This second stage calculation is performed in a commercial FPGA. The second pattern recognition step can proceed either via selection requirements on all fitted track candidates for by forming track seeds in the innermost pixel-strip modules and selecting stubs via a road-search algorithm.

Multiple processing engines are assigned to each tower, resulting in typical time multiplexing ratios of about 20. Efficient data dispatching for time and regional multiplexing is achieved by

using a full-mesh backplane ATCA platform, which provides high bandwidth, low latency, and flexible real time communication among processing nodes. One ATCA crate will be required per trigger tower. A custom ATCA board (Pulsar 2b) based on a Xilinx Virtex-7 FPGA has been prototyped with the goal of creating a scalable platform to demonstrate the AM-based approach. In addition, a prototype pattern recognition mezzanine card has been developed. In the final design, this mezzanine card hosts a powerful FPGA and a custom AM integrated circuit.

### 7.3.5 Correlator trigger

- Cavanaugh, 3pp

The Correlator Trigger receives input information from the Calorimeter Trigger, the Muon Trigger, and the Track Trigger, and performs object reconstruction and identification from the combined input information. The Correlator Trigger then transmits the lists of reconstructed objects to the Global Trigger, which applies the Trigger Menu. The current conceptual design of the Correlator Trigger provides parallel trigger paths for each of six primary types of trigger objects (electrons, photons, muons, taus, jets, sums) plus one *in situ* development path, and is assumed to be driven by the Track Trigger input bandwidth, since that drives the number of correlations that are possible between the different input information.

At design luminosity, the HL-LHC corresponds to an average of 200 minimum-bias interactions (pile-up) per beam crossing and the Correlator Trigger is designed to cope with a 2 GeV threshold on the transverse momenta of selected tracks. On average, each minimum-bias interaction is expected to result in about 30 charged tracks, of which approximately 1 track ( $\sim 3\%$ ) is expected to have a transverse momentum above 2 GeV. To conservatively cover any effects from Poisson tails in the pile-up distribution, long non-Gaussian tails in the transverse momentum distribution of charged tracks, or any other effects such as larger than expected backgrounds or better than expected LHC performance, a safety margin of a factor two is included. Currently, the track trigger primitive word is baselined to be about 100 bits long, which translates to  $\sim 35$  kb of information that is expected to be transmitted from the Track Trigger to the Correlator Trigger per LHC beam crossing. Assuming a similar amount of information is transmitted from the Calorimeter and Muon Triggers at an LHC beam crossing rate of 40 MHz, the total bandwidth which must be received and processed by the Correlator Trigger (for each trigger object path) is approximately 4.2 Tb/s.

The conceptual design of the correlator trigger is based on regional reconstruction followed by global calculations, leading to a two-layer processing system.

### 7.3.6 DAQ

- Mommsen, 1.5pp

The baseline Phase-II DAQ/HLT architecture is the same as for the currently implemented system, with a single level hardware trigger (L1) and a second level software trigger (HLT) on commercial processors. The main parameters of the DAQ/HLT systems for Phase-II, in comparison with the current system, are summarized in Table 7.3. The Phase-II upgrade results in entirely new sub-detectors and/or new readout systems. A first estimate of the event size amounts to roughly 4.1-4.5 MB in the range of PU 140-200 [57]. The design value for the event size to be handled by the DAQ is assigned accordingly, taking into account a small margin (10%).



Table 7.3: DAQ/HLT system parameters.

	LHC Run-I 7-8 TeV	LHC Phase-I upgr. 13 TeV	HL-LHC Phase-II upgr. 13 TeV	
Energy				
Peak Pile Up (Av./crossing)	35	50	140	200
Level-1 accept rate (maximum)	100 kHz	100 kHz	500 kHz	750 kHz
Event size (design value)	1 MB	1.5 MB	4.5 MB	5.0 MB
HLT accept rate	1 kHz	1 kHz	5 kHz	7.5 kHz
HLT computing power	0.21 MHS06	0.42 MHS06	5.0 MHS06	11 MHS06
Storage throughput (design value)	2 GB/s	3 GB/s	27 GB/s	42 GB/s

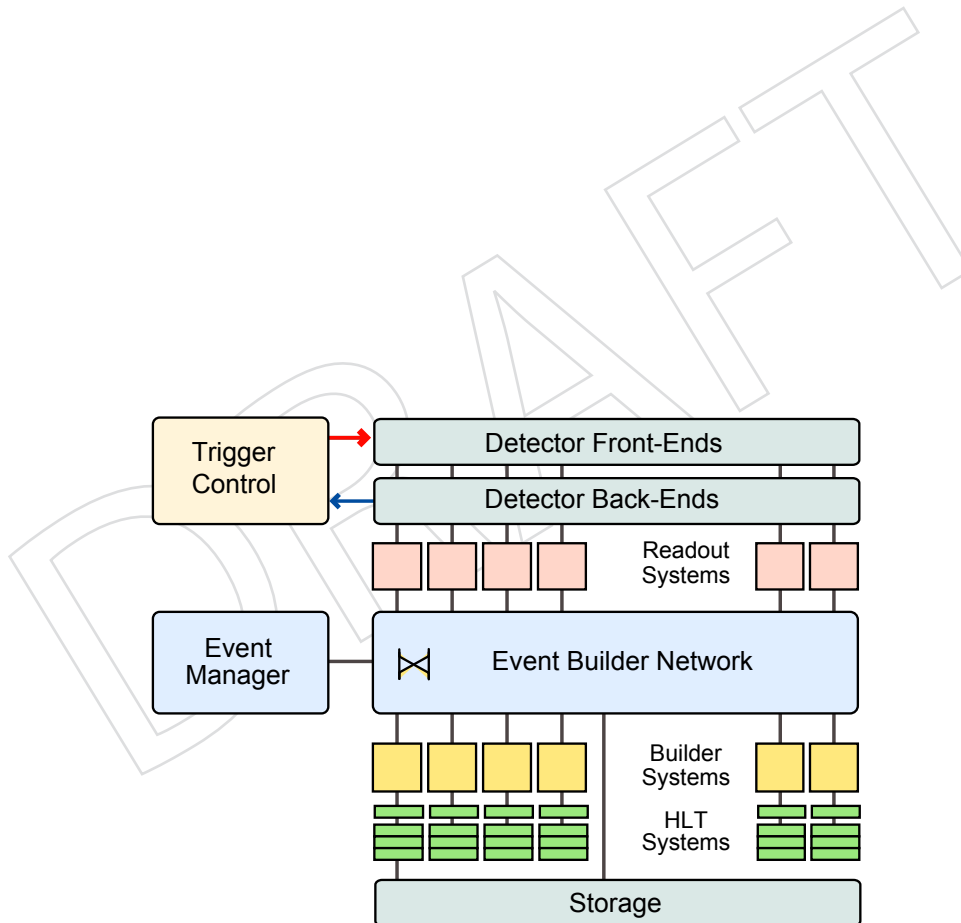


Figure 7.4: Schematic overview of the CMS DAQ system.



A schematic of the DAQ system is shown in Fig. 7.4. Read-out systems linked to the back-end electronics of the sub-detectors collect and buffer event fragments at the L1 rate. The read-out systems are connected via an event builder network to builder systems, which collect all event fragments belonging to the same L1 trigger and transmit the fully-built events to the nodes of the filter farm running the HLT. Hence, the HLT has access to the full detector information and is only limited by the available CPU resources and quality of online calibration. The event building process is steered by the Event Manager.

The local storage of the DAQ system for Run-II was designed for an effective throughput of 3 GB/s. This effective throughput is required for the storage of the main physics stream at the HLT output rate, as well as calibration and data quality monitoring streams. For Phase-II, the required throughput is determined by the event size and an HLT accept rate up to 7.5 kHz, whereas the rate of calibration and data quality monitoring streams is assumed to be similar as for Run-II.

Given the long lead time to Phase-II, it is premature to present details of a possible implementation that will be highly dependent on technology evolution. Depending on the details of the switching technology, experience has shown that event building can be achieved with an effective throughput 25% - 100% of the bandwidth of the switching network. The DAQ/HLT will be re-implemented taking advantage of progress in computing and networking industry in the next decade. The DAQ requirement in Table 7.3 correspond to an effective throughput of 30 Tbps at 200 PU for the event building network. Commercial networking equipment with a switching capacity of 50 Tbps and network interfaces operating at 40 Gbps (Ethernet) or 56 Gbps (Infiniband 4xFDR) are available today. The new DAQ system for Run-II is already based on these networking technologies, albeit with a smaller switching capacity. Hence, it appears feasible to anticipate a DAQ system in 2023 with the required throughput. As an example, a system with 800 data sources based on 100 Gbps links, and assuming  $\approx 38\%$  bandwidth efficiency, would provide the required throughput. This assignment includes sufficient reserve in the bandwidth of the DAQ links and switch ports to the event building network to accomodate that the data sources cannot be perfectly balanced in practice and that commercial network links have discrete transmission speeds.

The requirements on storage for Phase-II are a sustained throughput of  $\approx 40$  GB/s and a size of about 2 PB/s. This is an order of magnitude increase compared to the current system. Systems meeting these requirements are commercially available today, but expensive. However, the price-to-performance is expected to evolve significantly in the next decade.

## 7.4 Performance

- Savin/Acosta/Wittich/Cavanaugh 5pp

Describe expected performance for each subsystem with proposed algorithms.

At least one good test case per subsystem.

Correlator trigger will require several demonstrated benchmarks for track-matched objects.

specify L1 menu at the end satisfying total bandwidth constraints.

### 7.4.1 Calorimeter trigger with crystal granularity

Significantly improved spacial resolution of the trigger system allows to reconstruct electrons and photons in the calorimeter with high resolution both for position and energy measurement.

The algorithm of the electron reconstruction in the the trigger mimics closely the one used in the physics analyses, with number of simplifications required by trigger latency and architecture. The cluster is build around a seed crystal and includes information about isolation, pileup and possible bremsstrahlung corrections. The resolution of electron position reconstruction with respect to the expected "true" position in simulated events is shown in Fig. 7.5 left, the resolution of present trigger system is also shown. As expected the HL-LHC trigger will have significantly better position reconstruction then the current system. In Fig. 7.5 right the trigger rate for a single electron trigger for min-bias events is shown as a function of the trigger threshold. The new design improves the stand-alone trigger rate by factor 2-3 in the region above 15-20 GeV, that is the most important region for physics analyses.

A combination of calorimeter trigger with tracking information will allow to reduce trigger rate by a factor of 10 as shown in Fig. 7.6

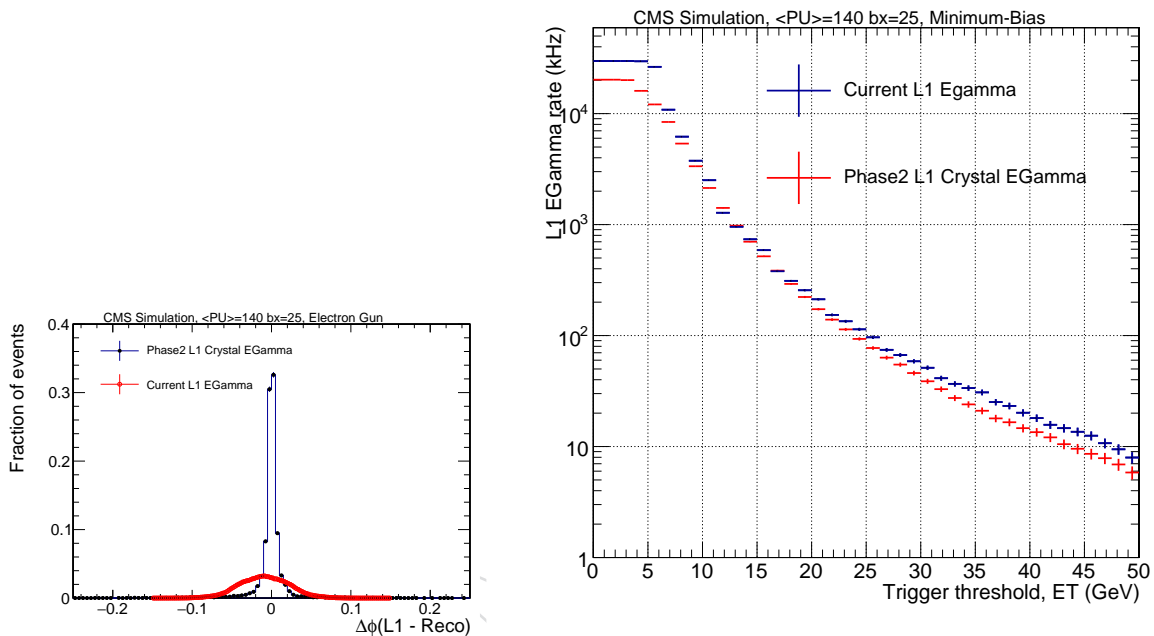


Figure 7.5: Left: The resolution of electron reconstructed position in the calorimeter with respect to the true expected position for Phase-II trigger reconstruction algorithm and for the current trigger; right: the expected rate for min-bias events using single electron trigger for Phase-II and existing trigger system.

## 7.5 Alternatives

- Berryhill 1-2pp

NSF vs. DOE scope swap for track trigger downselect. Brief description of NSF track trigger solution.

Pro/con of different tech/architecture choices being considered by us.

Consideration of alternative designs has been performed within constraints of upgrading the trigger system in an existing experiment. The system provides incorporation of alternative designs and architectures in the design itself so that as physics priorities and beam conditions

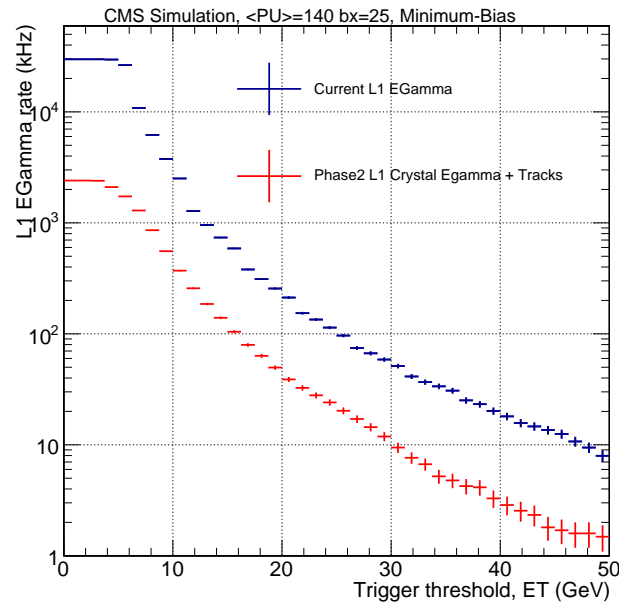


Figure 7.6: The expected rate for min-bias events using single electron calorimeter trigger matched to the track.

evolve, algorithms and trigger methodology can evolve as well due to the flexibility and programmability of the system. The design concept provides almost all available detector information at input of the trigger logic so that trigger decisions are not impacted by upstream selection of information. This enables changes in trigger design to have as wide a range of options as possible

## 7.6 Environmental Impact, Health, and Safety

- Berryhill 1pp

statement of ESH concerns for the project.

list of safety standards and procedures we plan to abide

Environmental Health and Safety:

Safety: follows procedures in CMS-doc-11587, FESHM. The L2 Manager for the Trigger Upgrade project is responsible for applying ISM to trigger upgrade, under direction of US CMS Project Management. The trigger modules are similar to others built before, of small size and no high voltage. They will be integrated into existing well-tested and long-term performing safety systems. All activities and personnel at CERN are regulated by CERN Safety Rules (e.g. safety training courses required of all personnel).

We apply a graded approach to risk. We apply appropriate the level of analysis, controls, and documentation commensurate with the potential to have an environmental, safety, health, radiological, or quality impact.

## 7.7 Risk

- Savin/Acosta/Wittich/Cavanaugh 5pp

about a 0.5 to 1.5 page per subsystem on risks and opportunities

technological risks for the hardware platforms: their mitigation and risk reduction

tech opportunities for nextgen platforms with smaller cost and more computing power.

risks/opps from scope of up/downstream components

risks/opps from change in scope with our international partners

algorithmic opportunities for nextgen platforms

algorithm risks from physics/detector performance unknowns at highest PU

In order to minimize risk in the project, the trigger upgrade is based upon common ATCA hardware platform and components (Xilinx FPGAs, multi-gigabit optical links) also used by other CMS systems. The system also has many handles to facilitate testing, including test-pattern injection, spy-buffer readout as well as test stands. Finally, risk is minimized by using many members from the Same team that built the existing trigger system and its Phase-1 upgrades and wrote its software and firmware. This team well understands the scope, requirements and interfaces.

A risk analysis of the trigger project has been performed. The list of risks includes the following with mitigation steps indicated.

- Senior Engineer becomes unavailable (Low Risk)
- Hire new engineer, subcontract to consulting firm, use FNAL engineer
- Funding is delayed (Low Risk)
- Commission with prototypes and/or fewer production boards
- Software or Firmware does not meet requirements (Low Risk)
- Hire extra expert effort to recover schedule and help personnel
- Boards are delayed (design, manufacture or testing) (Low Risk)
- Hire extra effort to speed up testing schedule
- Vendor non-performance (Low Risk)
- Acquire spending authority to use alternative vendors (while original funds are being unencumbered).
- Input or output electronics (non-trigger) delayed (Low Risk)
- Built in capabilities of trigger electronics provide signals for their own inputs outputs

## 7.8 Quality Assurance and Quality Control

- Berryhill 1pp

quality control procedures for

hardware

4573 optical connections

4574 memory R/W

4575 firmware

4576 online software

4577 offline software/emulation

4578 performance metrics with data and mc

4579 QA/QC: Testing and Validation

4580 Quality Assurance follows procedures in CMS-doc-11584. We regularly evaluate achievement  
4581 relative to performance requirements and appropriately validate or update performance re-  
4582 quirements and expectations to ensure quality. As mentioned earlier there is evaluation of the  
4583 trigger system electronics at CERN in the EIC and at point 5 in Local and Global Runs.

4584 Quality Control also follows procedures in CMS-doc-11584. Equipment inspections and ver-  
4585 ifications include Software code inspections, verifications, and validations; Design reviews;  
4586 Baseline change reviews; Work planning; and Self-assessments. All modules have hardware  
4587 identifiers which are tracked in a database logging QC data through all phases of construction,  
4588 installation, operation and repair.

4589 The plan for QA/QC involves full testing at the institute before shipping to CERN. All tests  
4590 are recorded (of all types) for individual boards in the database. The tests use and validate the  
4591 software and firmware test releases. After shipment and receipt at CERN there is acceptance  
4592 testing in the individual testing labs in the Electronics Integration Center (EIC) at CERN, where  
4593 the boards are retested to validate the institute test results. These tests use the same software  
4594 and firmware test releases as used in the institute testing. Following this individual board  
4595 testing, the testing moves on to Integration Testing in the EIC. The integration facility consists  
4596 of rows of racks with DAQ, Trigger, Central Clock, Crates of other subsystem electronics. This  
4597 facility tests operation of a vertical slice with electronics interfaces to other systems. These  
4598 tests use and validate the software and firmware commissioning release. After these tests are  
4599 completed, the electronics is then sent to point 5 and installed in the final rack locations with  
4600 final cables. The integration tests at the EIC are repeated to validate the system. The electronics  
4601 is then tested in Local Runs with data from the detector using test pulses and cosmic rays.  
4602 Finally, the electronics is included in Global Runs involving many other CMS subsystems using  
4603 cosmic rays and tests pulses. These full-scale tests use the full CMS DAQ/Trigger/Clocking  
4604 and the software and firmware commissioning release. After commissioning with the Global  
4605 Runs is completed, operations continue and the testing validates the software and firmware  
4606 initial operational release.

## 4607 7.9 Value Engineering

- 4608
- Berryhill 1pp

## 4609 7.10 R&D Towards Final Design

- 4610
- Berryhill 2pp

4611 outline of R&D program for

- 4612 high speed optical connections
- 4613 large memory modules
- 4614 Ultrascale+(+) mounted PCBs in ATCA form factor
- 4615 ZYNC SoC board control
- 4616 associative memory and other track trigger tech
- 4617 firmware algortihms
- 4618 system architecture

DRAFT



## Chapter 8

# CSC Muon Upgrade (DOE scope) - Editor Stan Durkin

## 8.1 Introduction

This report will update a plan to replace system FE boards that will fail when operating at HL LHC trigger rates. It will be showed that the new CMS Level 1 trigger latency and L1A rate for the HL-LHC combined with the increased luminosity will necessitate replacing on-chamber Cathode Front End boards (CFEBs) on the inner high rate chambers (ME2/1, ME3/1, and ME4/1). In concert new peripheral crate TTrigger Mother boards will be required to interface with the replaced CFEBs. Additionally, nearly all of the Anode Local Charge Track (ALCT) boards in the system will be replaced.

The present CSC system works well. It has dominated CMSs measurement of muons in the 1.5 to 2.5 eta region during all LHC CMS physics analyses. Once these rate problems are fixed, it will continue to work at the HL LHC. All electronics components in the system have been tested in a reactor or proton beam for radiation hardness[80][81][82]. All have shown little damage after irradiation with three times the expected integrated HL LHC radiation levels (10 Krad). The CSC chambers have been tested beyond HL LHC integrated wire charge and have shown little degradation[83]. The upgrade seeks to maintain the present CSC system by increasing the rate capabilities of the present boards. The physics capabilities of the device will remain as they have been for the last 10 years of data taking and analysis.

To discuss upgrading the system it is necessary to describe the basics of the present system. The CMS Cathode Strip Chambers (CSC) consist of 540 CSC modules in total situated in four stations between the 1.5 m magnet return yokes on each endcap[13][84]. Fig 8.1 displays the first station showing the trapezoidal CSC modules. Each CSC module consists of six gas layers, each layer having a plane of radial cathode strips and a plane of anode wires running perpendicular to the strips. The anode wires are ganged together to reduce the number of readout channels. There are a total of 266112 cathode channels and 210816 anode readout channels. The system covers the  $\eta$  region from roughly 1.5 to 2.4.

Shown in Fig 8.2 is a schematic of the CSC electronics system [85]. The three on-chamber boards of interest to this upgrade are the Low Voltage Distribution board (LVDB), the ALCT, and CFEB. The peripheral crate board of interest to this upgrade is the TMB.

- The LVMB regulates voltages required by the CFEBs, Anode Front end boards (AFEB), and ALCT boards.
- The ALCT board receives discriminated hits from the AFEBs. The ALCT creates trigger primitive Local Charge Tracks which are passed to the peripheral crate Trigger Motherboard(TMB). On receipt of an L1A it ships the wire-hit data in 25 nsec



and Northeastern(Optical fibers and cables). Our Russian collaborators are responsible for the LVDB. In terms of total cost 97% of the upgrade effort is within the US scope. Although not part of the scope of this upgrade other institutes will contribute heavily to the overall effort: U.C.Davis (radiation testing of electronics components), Florida(ALCT firmware and electronics system testing), Rice(computer code for electronics communications, U.C.S.B. (electronics system testing), and U.C. Riverside (electronics system testing).

## 8.2 Requirements

In this section we reiterate the arguments given for building CSC chambers given in the CMS Muon Project technical design report published in 1997[84].

Cathode Strip chambers were chosen because their intrinsic resolution can be as good as 50  $\mu\text{m}$ . They also operate in large magnetic fields without significant deterioration of their performance. The intrinsic resolution is defined by the signal-to-noise ratio, which we require to be 100:1. This ratio is met by attaching each cathode strip to the Buckeye amplifier ASIC coupled to a precision 12-bit ADC. Each CSC chamber was built with 6 planes providing redundancy and guaranteeing a precision measurement of both the position and angle of a measured muon in the bending plane. Ganged wires provide the precision timing for the resulting track segment. The ALCT amplifier (25 nsec peak)/discriminator ASIC was built to meet these specifications.

Chambers are arranged to form disks, called stations (ME1, ME2, ME3, and ME4). There are a total of 540 chambers in the system. The station ME1 has three rings (ME1/1, ME1/2, and ME1/3), while the other three stations have two rings (MEx/1 and MEx/2). All but ME1/3 chambers overlap in eta and thus form seamless rings. The cracks between the chambers are not projective, and thus the coverage, defined as at least three chambers on a muon path is close to 100%. The high redundancy of the CSCs is a central feature of the design and is responsible for the robustness of the system in muon triggering and reconstruction.

This proposed upgrade is driven by the rate limitations of the CSC electronics built from 1994 to 2004. The data queue designs inside the CFEB and ALCT boards will fail at HL LHC data rates in the central chambers. Additionally the DMBs serving these same boards have an insufficient bandwidth to transfer the data to the CSC FED system. These are the only fixes that need to be made to the system. The information provided by the detector at the HL LHC will be essentially the same as it is now. Cathode trigger primitives will still be based on the comparator ASIC. The cathode digitized signals will still be amplified and shaped by the Buckeye amplifier shaper ASIC [85]. The precision of the 12 bit ADCs is nearly identical. The Anode trigger will not change, and the anode data will be amplified and discriminated on the same Anode Front End Boards.

The CSC system was designed for precision measurement, and to have sufficient redundancy to work in noisy track-reconstruction conditions. These design choices are even more important in the HL-LHC environment where the chambers get many extra hits from neutrons from pileup events.

## 8.3 Proposed design

### 8.3.1 CFEB Rate Problems

The Cathode Front-End boards (CFEB) are presently used on all CSC chambers except ME1/1 (upgraded 2013-2014). Designed and built in 1996-2002, the CFEBs use analog charge storage within custom Switched Capacitor Array (SCA) chips (ASICs) that sample at 50 ns intervals and contain a depth of 96 cells (6 events worth of data) during Level 1 trigger latency and 26  $\mu\text{sec}$  of charge digitization (Wilkinson ADCs). At HL-LHC increased Level 1 latency (12.5  $\mu\text{sec}$ ) and Level 1 rate (up to 750 kHz) the 108 inner CSC chambers (ME2/1, ME3/1, and ME4/1) will overwrite capacitor cells in the SCAs causing large readout inefficiency. Trigger rates in outer CSC chambers are lower than those in the inner rings by a factor of 5 or more with the exception of ME4/2 which has about thirty percent of ME4/1 the rate.

The problems outlined above can be overcome with selective replacement of CSC electronics. The SCA overwriting problem can be addressed by replacing the 540 CFEBs in ME2/1, ME3/1, ME4/1 with the Digital CFEBs (DCFEBs) that were developed for the ME1/1 chambers and installed in that region during LS1. To accompany the DCFEBs, 108 Trigger Motherboards (TMBs) would be replaced with Optical TMBs (OTMBs). The OTMBs were also part of the LS1 upgrade of ME1/1. The 108 ALCT mezzanine boards would be replaced with new Optical ALCTs (OALCT) mezzanine boards.

The current shutdown schedules provided by CMS technical coordination require that the CFEB, TMB, and ALCT board replacement will take place in LS2. Additional off-chamber electronics boards are also to be upgraded in LS3, and these are not part of the DOE funding request, but are part the NSF MREFC proposal for the ATLAS and CMS upgrades.

Shown in Fig 8.3 is the fractional event loss versus luminosity for the CFEB electronics in ME2/1, ME3/1, and ME4/1 chambers. Data losses increase dramatically as luminosity increases. At the present Level 1 Latency (3.4  $\mu\text{sec}$ ) and Level 1 rate (100 KHz) the CFEBs would work well at the HL-LHC. Using the HL-LHC trigger design of 12.5  $\mu\text{sec}$  Level 1 Latency and 750 KHz Level 1 rate, the inner ME2/1, ME3/1, and ME4/1 chambers will sustain large data losses ( $> 10\%$ ), and thus their CFEB boards clearly need to be replaced.

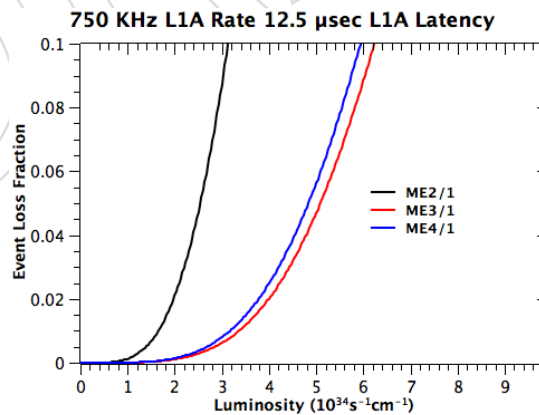


Figure 8.3: Event Loss Fraction vs Luminosity (the ultimate HL LHC luminosity is  $7.5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ ).

These CFEB event loss rates depend on the chamber preLCT rate (12.5  $\mu\text{sec}$  L1A latency SCA storage) and the LCT\*L1A rates (26  $\mu\text{sec}$  SCA digitization queue). Here the preLCT is a faster but less precise version of the TMB LCT optimized to decrease SCA usage. The rate extrapolation

tion is done starting with 2015 data. The preLCT rates are scaled linearly with luminosity to the HL LHC luminosity. The LCT\*L1A rates are scaled linearly by luminosity and again linearly with the L1A accept rate to the HL LHC luminosity. These estimates are not conservative. No HL LHC Rate uncertainty has been included in the calculation.

In December of 2015 we measured the CFEB overflow rates by pulsing a CFEB at the rates calculated from this extrapolation. For this measurement, firmware for the CFEB and DMB was modified for HL LHC running condition (12.5  $\mu\text{sec}$  latency, a 3 bx L1A\*LCT window, and optimized for 750 KHz L1A accept). Shown in Fig 8.4 is a comparison between our pulser tests and the statistical model. The measured loss rates fit a simple statistical model (poisson+5 event storage queue) well (within 7%). Thus our extrapolations are all based on measurements of the actual CFEB hardware.

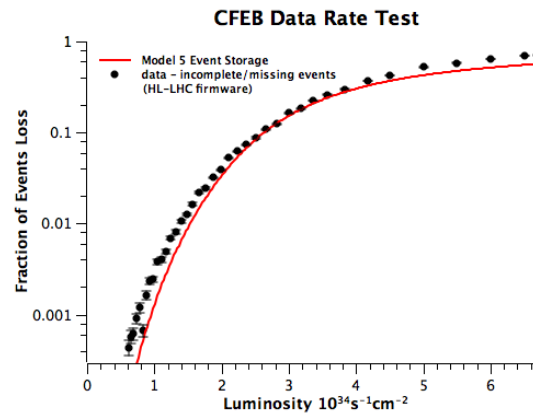


Figure 8.4: Measured CFEB (with HL LHC firmware) rate loss vs statistical model.

The TMB boards have trigger counters imbedded in the firmware to keep track of relevant CSC trigger rates. These are read out periodically during LHC running and stored to disk. Given our extremely good understanding of the CFEB rate limitations, our largest uncertainty are these trigger rates. The preLCT rates, based on TMB tracking patterns, have been shown to scale linearly, with nearly constant preLCT rate/luminosity in a given chamber type during the 2015-2016 LHC running. In contrast the LCT\*L1A accept rates depend on Level 1 Trigger tables. The Trigger tables control what event topologies are triggered and sets their prescaling rates. They vary run to run when the luminosity is low (special triggers), and become reasonably constant at higher luminosities. Shown in Fig 8.5 is the LCT\*L1A rates per luminosity for the inner chambers. The 2015 data are seen to asymptotically approach a single value at higher luminosities. This makes sense because at low luminosity the trigger tables turn on nonrestrictive triggers, but at higher luminosities physics triggers dominate. The difference between the 2011 data asymptote and that of 2015 is attributed to the LHC 8 TeV and 13 TeV center of mass energies in 2011 and 2015 respectively. The 13 TeV trigger rates per luminosity were roughly 25% higher than those at 8 TeV.

### 8.3.2 DCFEBBoards

The limitations of switched capacitor arrays have long been understood by the CSC group. The CFEBs were designed and constructed between 1996 and 2004. The best storage available at the time was switched capacitor arrays. The DCFEB was designed for the LS1 ME1/1 upgrade. The idea was to replace the switched capacitor arrays and Wilkinson ADCs with Flash ADCs followed by a digital pipeline inside the Virtex 6 FPGA. This yields a nearly dead-timeless system. The output rate is limited only by the 3.2 Gbps optical output bandwidth and event



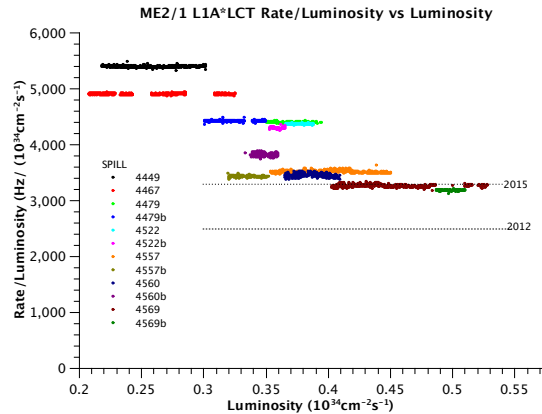


Figure 8.5: L1A\*LCT rate/luminosity vs luminosity. The L1A\*LCT rate is per chamber. The dashed lines represent the asymptotic rate/luminosity for 2012 and 2015 data.

storage (10 event storage on a Virtex 6 uses negligible chip resources). The board described below has been installed and working on ME1/1 for LHC data taking in 2015, and 2016. The plan is to build 540 boards plus spares to replace CFEBs in the inner CSC chambers (ME2/1, ME3/1, and ME4/1) during LS2.

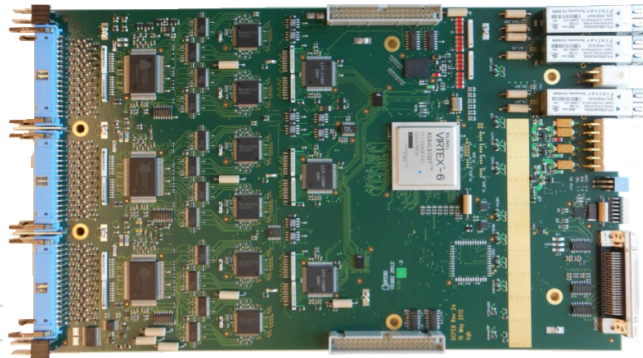


Figure 8.6: DCFEb board

The DCFEbS amplify, shape, and digitize charge signals from the cathode strips. A custom low-noise amplifier/shaper ASIC is used to attain a 1% charge measurement precision. The 96 amplified channels are converted from single ended to differential signals in a fully differential amplifier. Each of the 96 channels is then digitized by a fast flash hybrid 12-bit ADC at 20 MHz. The ADC data is then fed into a Xilinx Virtex-6 FPGA which pipelines the data during the L1A latency. If an L1A\*ALCT\*CLCT coincidence is seen, the data is transferred to the OTMB via a 3.2 gigabit optical fiber. If no coincidence is seen, the data passes unused out of the pipeline.

The DCFEb is an upgrade of the original CFEB board. It would be nearly impossible to improve on the original comparator and amplifier/shaper ASICs. Thus the analog portion of the board is nearly identical to the original CFEB boards. Spark protection on the inputs is identical. The Buckeye amplifier/shaper ASIC3 is used on both boards. The same comparator ASIC is also used on both boards, and the treatment of the analog signals up to trigger primitive outputs is also identical.

At this point the CFEB and DCFEb designs part ways. The CFEB made use of switched capac-



itor arrays for analog storage of events with 96 capacitors per channel. On receipt of a preLCT the data is stored on a capacitor during the L1A latency. Once an L1A\*LCT coincidence is seen, the CFEBs use 6 Wilkinson type ADCs to digitize the data and pass it to the DMBs via channel links through copper Skewclear cables. Additionally, the CFEBs passes the trigger primitives to the TMB also via channel links through copper Skewclear cables.

The major limitation of the DCFEB system is the bandwidth of the fiber link from the fixed queue. Shown in Fig 8.7 is the event loss versus luminosity for a ME2/1 CFEB and DCFEB. It is seen that the DCFEB outperforms the CFEB by a factor of more than ten in the rate.

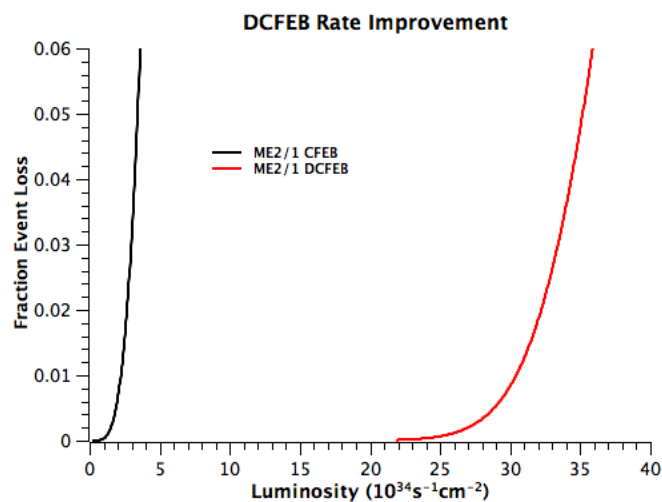


Figure 8.7: ME2/1 CFEB and DCFEB rate loss curves.

### 8.3.3 OTMB Boards

The DCFEB trigger primitive data is passed to the peripheral crate OTMB via a 3.2 Gbps optical fiber. Since the present TMB is designed to work with CFEBs, it has no optical receivers to accommodate the optical data link from the DCFEBs. This necessitates building 72 OTMB mezzanine boards and 72 OTMB main boards in LS2 for all ME2/1, ME3/1, and ME4/1 chambers.

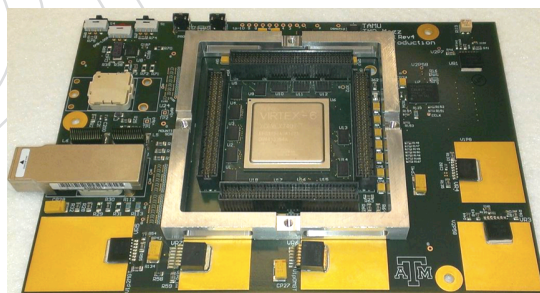


Figure 8.8: A OTMB mezzanine board. The snap-12 Optical transceiver can be seen on the left.

OTMB boards were designed and built in LS1 on the ME1/1 upgrade. The mezzanine boards utilize a Snap 12 connector featuring 12 optical input transceivers shown in Fig 8.8. This mezzanine board plugs into a VME OTMB baseboard. The boards also have a relatively modern Xilinx Virtex-6 FPGA with built-in high speed serial transceivers. The use of an upgraded FPGA will provide the additional advantage of increased logic resources to improve the trigger algorithms in this high rate region.

### 8.3.4 ALCT Mezzanine Boards

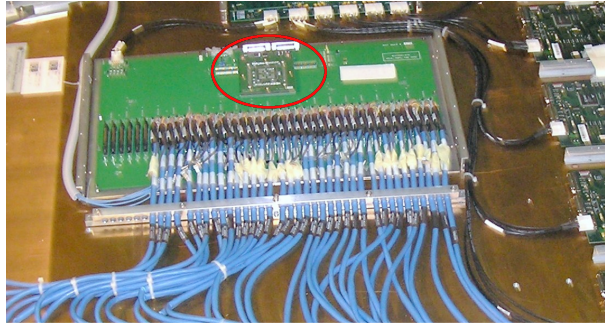


Figure 8.9: ALCT main board with ALCT mezzanine board (red circle)

The ALCT boards have a rather fundamental flaw with respect to the HL LHC L1A latency. The wire-hit data is transferred to the TMB board on receipt of an L1A. The data is stored in a pipeline within the virtex-E FPGA while waiting for the L1A. With our present latency of 3.4  $\mu\text{sec}$ , there is enough BRAM available for this pipeline. When the latency is increased to 12.5  $\mu\text{sec}$ , there is not enough BRAM resources to create a pipeline of sufficient depth. Therefore, the FPGAs on all the ALCT mezzanine boards need to be replaced, with the exception of ME1/1 and ME4/1 chambers, which were equipped with Spartan-6 FPGAs during LS1.

There is also another fundamental rate problem with the ALCT data readout. The existing boards transfer 16 time-bins (25 nsec) worth of data to the TMB board (through the rear transition board) through copper Skewclear cable. At HL LHC L1A rates, in the high rate chambers (ME1/1, ME2/1, ME3/1, ME4/1) the bandwidth of the copper connection is not enough to handle more than one time-bin worth of data. The solution to this limitation is relatively simple and inexpensive. The Spartan-6 FPGA is equipped to communicate with a 3.2 Gbps optical transceiver. By adding this transceiver to the mezzanine board we can easily transfer 16 time-bins worth of wire data to the OTMB snap12 connector, or directly to the new ODMB when they are installed in LS3.

### 8.3.5 Links

Two 3.2 Gbps optical fibers will be installed in LS2 from each DCFEB to the Peripheral crates. The first will carry the trigger primitive data to the OTMB mezzanine boards Snap-12 connector. The second will carry digitized ADC strip data. It will not be active until LS3 when it will be connect to the ODMB. Existing Skewclear copper cables will carry clock signals, JTAG communication lines, and ADC strip data from the DCFEBs to the DMBs until LS3 when the optical link becomes active (see Staging section below).

### 8.3.6 LVDB

The voltages required by the DCFEBs are different from the CFEs. To solve this problem, 108 new LVDBs will be built to replace the present LVDB boards mounted on the inner CSC Chambers.

### 8.3.7 Staging

The boards described in this section (DCFEB, LVDB, ALCT mezzanine, OTMB) will be installed in LS2. In LS3 the ME1/1, ME2/1, ME3/1, and ME4/1 both the DMB and ODMB boards will be replaced with new higher rate ODMB boards, and the entire Front End Driver (FED) system

will be replaced. Waiting to design these higher data rate systems makes sense given that optical links and FPGAs are improving in bandwidth year by year.

This means that the ME2/1, ME3/1, and ME4/1 DCFEBs will have to communicate with the old DMBs between LS2 and LS3. The DCFEB was designed with two paths for output data, one optical and the other copper on the Skewclear cable used for communications between CFEs and DMBs. We have tested and have shown that the DCFEBs will communicate with both DMBs and optical DMBs.

### 8.3.8 Infrastructure

The new DCFEBs consume more power than the old CFEs. To solve this, multiple new Maraton supplies will have to be added to the CSC system. With new Maratons, new junction boxes must be built for redistributing the power.

## 8.4 Performance

As stated previously the information output of the DCFEB data after the upgrade is identical to the information before. All that has changed is the replacement of the switched capacitor arrays with FLASH ADCs and a digital pipeline. The data resolution, format, and unpacking remain exactly as before. All physics analyses done with muons since 2007 have used the CSC to find and measure muons in the region  $1.5 < |\eta| < 2.5$ . Literally hundreds of approved plots are available for online monitoring CSC performance [86] detailing efficiencies, resolution, and timing among other details. Little would be gained by summarizing the details here. The CSC system has worked extremely well and at any time  $>2\%$  of channels on the detector are not working.

Performance of the CSCs at an HL LHC pileup of 200 with large photon/neutron backgrounds has been and will continue to be studied. Monte Carlo simulations of Higgs decaying into four leptons have been studied. Backgrounds included pileup but no photon/neutron thermal gas. Not surprisingly with the CSC redundancy, no tracks were lost, and there was little degradation of measurement of the Higgs invariant mass.

The GIF++ facility has been used to understand HL LHC photon/neutron backgrounds. The facilities have a tunable Cs-137 source and a 100 GeV muon beam. Figure 8.10 shows cathode strip measurements for a typical GIF++ event. The Cs source has been set to mimic the HL LHC photon/neutron backgrounds. A 100 GeV muon is seen passing through all six planes at the center of the plot. There is clearly a large amount of noise from the Cs photons. They typically have very large charge deposits.

We designed the electronics to output precision ADC information for just this reason. With this information we can implement filters to suppress such backgrounds. The precision ADC data is most important in very noisy events at the HL LHC.

Running with the muon beam and a sizeable HL LHC Cs photon background, a recent CSC study has shown a small 2% loss in the segment finding efficiency in ME1/1b and ME2/1 chambers. Resolution in the bending plane is increased on ME2/1 from a nominal  $115 \mu\text{m}$  with low background rate to  $203 \mu\text{m}$  at the HL LHC setting. Similarly, resolution in the bending plane is increased on ME1/1b from a nominal  $58 \mu\text{m}$  with low background rate to  $80 \mu\text{m}$  at the HL LHC setting. The results are very encouraging.

At the present time we have no mechanism to study the system under full HL LHC conditions with a pileup of 200 and photon/neutron backgrounds. We will study the high pileup run

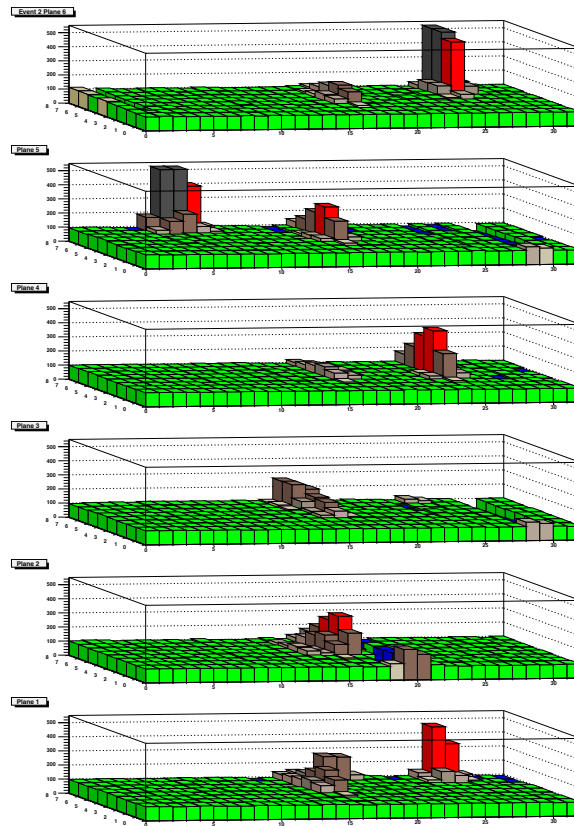


Figure 8.10: Typical GIF++ plot of a muon passing through the 6 planes of the ME1/1b Chamber. The front axis corresponds to 32 wire strips. Eight time samples are shown on the axis into the page, and the vertical axis is ADC counts. Colors emphasis ADC counts (measured charge) in a given time sample. Minimum ionizing particles have typically light brown level deposits, while red and black correspond to greater than 3 and 5 times minimum ionizing respectively.

with HL LHC-like taken in late October 2016 to improve our understanding of the HL LHC conditions.

## 8.5 Alternatives

As stated before the present CSC system works well. It is a huge system. The on chamber electronics alone consists of 2,772 CFEBs/DCFEBs, 540 ALCs, and 540 LVDBBs. It would be prohibitively expensive to replace the entire system. The upgrade is cost effective. By replacing 25% of the present CFEBs in the system, 90% of the digitized data will pass through DCFEB boards. The remaining 75% of the boards in the system will continue to work well at the HL LHC. In short there is no other alternative that would maintain the physics capabilities of the present system.

## 8.6 Environmental Impact, Health, and Safety

The environmental impact from the boards is expected to be extremely small. This is very standard electronics. US companies manufacture the PC boards. All part are standard off the shelf items including FPGAs, connectors, clock, and flash ADC, among others. The two ASIC

were manufactured by industry. They are similar to boards on modern computers and TVs. As such their environmental impact is similar to these item. The cables are all non-halogen per CERN regulations.

On the safety side supply to the boards is 8 Volts with fairly high current. As such there is no danger of electrocution. Fire risks are present, as with all electronics systems. As such each board has fuses on the various voltages supplied to the boards. Computers also monitor the current, temperature, and voltage on the boards continuously when powered. Once again non-halogen cables are required and have past a CERN flammability test.

Radiation safety is important for individuals maintaining the electronics on the CMS detector, as well as those working at test facilities at CERN. Radiation areas include the GIF++ test facility, and the detector hall at Point 5. All CERN regulations, radiation training, and dosimeter requirements are strictly followed by all of our personnel. Databases for boards on the detector are maintained and tracked by CERN. Electronic boards replaced on the CMS detector are put in a holding area until radiation levels have dropped below safety standards. Once released they can be safely repair and stored as spares.

Radiation safety is also strictly followed at reactors and proton beams used to irradiate component for radhard testing. Facilities used are both in the United States and at CERN. The radiation safety rules are specific to the given facilities. Boards are not released back to the CSC group until radiation levels have dropped to below safety standards.

There are no health problems associated with the proposed electronics.

In the multiple electronics reviews both by US CMS and also by CERN, environmental impact and safety are part of the discussion. All electronics on the detector have been reviewed and comply with CERN and European regulations.

## 8.7 Risk

Currently our baseline is to build exact copies of the DCFEBs, OTMBs, and ALCTs produced in LS1 production. The layouts exist. Nearly all of the components, although past state of the art, are still available. The ALCT mezzanine board will need to add an optical transceiver but this is considered a minor change. The boards meet HL LHC performance and we have been taking data with all three boards over the last two years. We still reserve as a possibility implementing a newer FPGA on the new DCFEBs. This would have the advantage of using less power, and having a more modern FPGA with more options. This is not a major change in the existing DCFEB board. The fact that we already have a working system goes a long way toward mitigating most risks.

We know once built and installed the electronics will perform well. The remaining risks involved are in the production schedule. Procuring parts could be delayed. Production of the PC boards, or stuffing could be delayed. Having the funds to start production well before the boards are needed in LS2 can control these risks.

## 8.8 Quality Assurance and Quality Control

The DCFEBs, ALCTs, OTMBs, and LVDB are situated in moderate radiation areas (10 kRads HL-LHC integrated radiation) with strong magnetic fields. Every electronic component on these boards has or will be irradiated in a proton beam, or reactor to test its radiation hardness.



They are tested to 30 Krad. FPGAs, ADCs, and transceivers are also tested to measure their Single Event Upset (SEU) cross-sections.

There are two ASICs on the DCFEB Board, the Buckeye Amplifier/Shaper ASIC and the Comparator ASIC. These chips are individually functionally bench tested. A few chips are also radiation tested. There are enough Buckeye ASICs left over from our year 2000 production. Additional Comparator ASICs were ordered in November 2016, and will be thoroughly tested when the chips arrive.

We also test electronics boards during production. For example the DCFEBs in 2012 on arrival were subjected to a 20 computer controlled tests in a test stand at Ohio State to assure the boards were functional. If the boards had problems they were repaired and retested. The results of the tests were archived for future reference. There are test stands at UCLA and TAMU for the ALCT and OTMB boards as well.

The boards are shipped to CERN in relatively small numbers to mitigate any potential loss. Once the boards arrive at CERN they are retested in the test stand at building 904. Lastly, during refurbishing the old boards are removed from chambers and the new boards installed. The boards then undergo a full systems test before the chamber is remounted on the detector.

## 8.9 Value Engineering

The DCFEB, OTMB, and ALCT mezzanine board designs already exist from the 2012. We reserve as a possibility changing the FPGAs on these boards to increase functionality, decrease power usage, or decrease costs. In the case of the DCFEB the choices are rather constrained. FPGA IO has to be large given that there are 96 differential pairs from the flash ADCs and 48 signals from the comparators. The obvious value engineering that applies here is waiting for the latest technology. This is especially true for FPGA and fiber links. We are forced to install the on-chamber electronics in LS2, but it is advantageous to put off the design of the ODMB and FED electronics until LS3 to use the newer, faster, and improved FPGAs and optical links. This improvement in FPGA between our CFEB Virtex FPGA and the DCFEB Virtex-6 FPGA is nothing short of astounding and the price was not vastly different. Chip temperature and current modeling, SEU correction[87], huge block rams, and 3.2 Gbps serial transceivers were all new to the Virtex-6. The large block rams were absolutely necessary in creating the digital pipeline essential to the workings of the DCFEB.

## 8.10 R&D Towards Final Design

The designs of all three electronics boards we intend to build for LS2 have to be revisited. In the case of the DCFEBs it may be prudent to replace the Virtex 6 FPGA in the 2012 board design with a newer less power hungry FPGA. This would potentially reduce power, lower costs, and improve FPGA performance.



## Chapter 9

# CSC Muon Upgrade (non DOE scope) - Editor Alexei Safonov

## 9.1 Introduction

In the Long Shutdown 3 (LS-3), the CMS muon system will undergo several upgrades to ensure continued good performance of muon triggering and identification in the High Luminosity LHC regime. The aim of the upgrades is to maintain good performance of the existing detectors in the new environment and to improve CMS muon triggering capabilities to meet the requirements set by the physics program of the HL-LHC operation. Current CMS muon system brings together three technology choices: the Drift Tube (DT) chambers in the central region (the “barrel”)  $|\eta| < 1.1$ , the Cathode Strip Chambers (CSC) in the forward region (the “endcap”)  $1.1 < |\eta| < 2.4$ , and the Resistive Plate Chambers (RPCs) both in the barrel and in the endcap extending to  $\eta = 1.6$ . The information provided by the Muon detectors forms the basis of muon reconstruction, identification and triggering at CMS, including a standalone momentum measurement required by the Level-1 trigger. Both the barrel and endcap muon detectors are arranged in four stations of chambers separated by steel absorber. The redundancy is necessary to ensure a good measurement of the muon momentum, high reconstruction and identification efficiency and low rate of misidentifications.

The increased data and trigger rates require partial replacement of the data acquisition and trigger electronics for all three existing systems (DT, CSC, and RPC). In addition, CMS plans to install new muon detectors in the forward part of the endcap region that has not been instrumented with the RPC detectors at the time of original construction due to too high incident particle rates for then-available RPC technology. For the LS-2, CMS will install the first of these new detectors, the GE1/1 muon detector utilizing a relatively recent Gaseous Electron Multiplication (GEM) technology. The decision for early installation of GE1/1 has been driven by the desire to take advantage of the the improvement in trigger performance achievable with the new detector. In the LS-3, CMS plans to install GE2/1, a second GEM-based detector, so that the two stations closest to the interaction point in the most-forward region of the CMS muon system will be instrumented with GEM detectors. The two far stations will be instrumented with the improved RPC detectors forming stations RE3/1 and RE4/1. The new detectors are required to maintain an efficient muon triggering in the region  $|\eta| > 1.6$ , e.g. integrating the data from the new GEM detectors with the CSC data has been shown to lead to a strong improvement in trigger performance (see below). In addition, CMS plans to install a new very forward muon detector ME0 extending muon offline coverage from  $|\eta| = 2.4$  to  $|\eta| \simeq 2.9$ .

The will play several key roles in the upgrades to be installed in LS-3 focusing on the areas of its historical strength and responsibilities in CMS. First, in LS-3 the will perform an upgrade of the off-chamber elements of the CSC data acquisition system to increase its throughput to

match the HL-LHC requirements. While the LS-3 CSC upgrade addresses a specific problem of insufficient throughput of the CSC off-chamber electronics, it is complementary with the upgrade of the on-chamber and OTMB electronics planned for the LS-2, and with its completion will enable efficient and robust continuous operation of the CSC detector. Second, the will design and build the on-chamber Optohybrid boards and the backend electronics for GE2/1 and ME0 detectors. These elements are essential in enabling the integrated Level-1 triggering in the forward region, in which the trigger data from the nearby GEM and CSC chambers is combined to measure local muon direction. The latter has been shown as a powerful tool in rejecting backgrounds in the Level-1 trigger.

## 9.2 Requirements

The requirements for the two sub-projects, the CSC off-chamber electronics and the GEM trigger/DAQ system, are somewhat different and are driven by the respective goals of each upgrade.

For the CSC off-chamber electronics upgrades, the requirements are as follows:

- Deliver 180 ODMB boards (plus spares) complete with the firmware for the ODMB FPGA and online software for operation of the ODMB boards
- Deliver 120 fiber bundles of 12 fibers each complete with optical fanouts to connect the ODMB boards in the existing 60 CSC peripheral crates to the new FED system
- Deliver 2 crates (+1 spare) and 12 processing modules (+3 spares) to build the CSC FED system complete with required firmware and online software
- All electronics components are required to meet safety and environmental requirements set by CERN and be able to sustain radiation exposure of the entire HL-LHC running period
- The ODMB and FED system has to be able to operate new links at up to 10 Gbps data transmission rate
- The FED system design has to be compatible with the central CMS DAQ system

For the GE2/1 and ME0 DAQ/Trigger electronics systems, the requirements are as follows:

- Deliver 144 GE2/1 Optohybrid (OH) boards (plus spares) complete with the firmware for the OH FPGA and online software for operation of the OH boards
- Deliver 2 crates (+1 spare), 10 processing modules (+1 spares) and 2 DAQ link cards for the GE2/1 backend electronics system complete with required firmware and online software
- Deliver 648 bi-directional optical links and 288 uni-directional optical links for connecting GE2/1 OH boards with the GE2/1 backend system
- Deliver firmware and online software required to operate the GE2/1 DAQ system and provide necessary data readout and triggering capabilities, including the link with the CSC OTMB board and required firmware modifications for the CSC OTMB board
- Deliver 216 ME0 Optohybrid (OH) boards (plus spares) complete with the firmware for the OH FPGA and online software for operation of the OH boards
- Deliver 2 crates, 16 processing modules (+1 spare) and 2 DAQ link cards (+1 spare) for the ME0 backend electronics system complete with required firmware and online

software

- Deliver 1080 (+ 108 spares) bi-directional optical links and 432 (+ 44 spares) uni-directional optical links for connecting ME0 OH boards with the ME0 backend system
- Deliver firmware and online software required to operate the ME0 DAQ system and provide necessary data readout and triggering capabilities
- All electronics components are required to meet safety and environmental requirements set by CERN and be able to sustain radiation exposure of the entire HL-LHC running period
- The OH and backend system have to be able to operate bidirectional links at rates compatible with the radiation hard GBTX hardware
- The GE2/1 and ME0 backend system designs have to be compatible with the central CMS DAQ system

## 9.3 Proposed design

### 9.3.0.0.1 CSC Off-Chamber DAQ Electronics and the FED System

The increase in the instantaneous luminosity coupled with the increased in Level-1 trigger latency and the accept rate will lead to a large increase in the data rates. One specific issue for the CSC data acquisition (DAQ) system is the insufficient bandwidth of the existing optical link between the CSC electronics in the peripheral crates (PC) and the Front End Driver (FED) system. The PCs are located in the racks attached to the sides of the CMS detector and contain electronics that concentrates and processes data received from the front-end electronics located on the CSC chambers, each PC serving a fraction of the chambers in one of the four stations of the CSC system. The FED system is responsible for concentrating the data from the entire CSC system and communicating it to the central CMS DAQ system via optical links. The FED system is located in the counting room. At the HL-LHC luminosity and the increased L1 trigger accept rate, the data rate in the PC-FED link significantly exceeds the bandwidth of the existing optical link for CSC stations ME1/1, ME2/1, ME3/1 and ME4/1, which are located in the most forward part of the system. Resolving this problem requires a new, higher speed link as well as a replacement of the Optical Data MotherBoard (ODMB) cards in the PC that send the data over the link as well as a replacement of the legacy FED system. Figure ?? provides an illustration of the layout of the CSC readout system and its main components.

Majority of the ODMB boards have been built during the original detector construction, but 72 of them (about 15% of the entire system) has been replaced in 2012 as part of the upgrade of the readout for station ME1/1, which is the closest to the interaction point part of the most forward part of the system. The plan is that the replacement version of the ODMB card will closely follow the design of the 2012 ME1/1 ODMB board, but will add features accommodating for the higher bandwidth of the new optical link for the data to be sent to the FED system. The new optical links to connect the ODMB cards to the FED system will use commercial off-the-shelf technology. Given the existing ME1/1 ODMB design, availability of the engineers who developed it, and the RD studies performed in 2015 and 2016 to verify compatibility of the ME1/1 ODMB with the planned electronics configuration for HL-LHC, this is a well understood project that requires only minimal additional RD related to selecting suitable replacements for parts that can become obsolete and building and testing a preliminary prototype of the board.

The FED system has been built during the original detector construction and is comprised of custom Detector Dependent Unit (DDU) cards arranged in four 9U VME crates. The increase

in the speed of the new links, concerns about the compatibility of the FED system with the significant technology upgrades on the central CMS DAQ system, which FED sends its data to, and the age of the technology of the original system, makes partial replacement of the components (40% of the system would need replacement) of the existing FED system technologically difficult to accomplish. Instead, we plan to build a compact replacement FED system based on either a  $\mu$ TCA or ATCA standard widely adopted at CMS and use one of the “standard” multi-purpose CMS boards. The most likely choices would be the next generations of the TCP7 or MTF7 high-power high-throughput boards used for the Phase-1 trigger upgrade. Use of one of the existing boards avoids costs associated with the design of a dedicated board providing significant cost savings and reduces personnel needs for electronics design and development to focus on the firmware development and integration of the new system with the existing elements of the CSC readout and the future central DAQ system.

#### 9.3.0.0.2 GEM DAQ and Trigger Electronics

The new GEM detectors GE1/1 and GE2/1 planned as part of the CMS upgrade are designed as “additional layers” for the existing CSC chambers, effectively forming large lever-arm tracking detectors within each station capable of measuring the direction of the muon track in addition to the track spatial position. Forward muon stations 1 and 2 are the optimal place for such detectors as the muon direction in these stations is strongly correlated with the muon momentum (in stations 3 and 4, in contrast, the correlation is reduced due to the trajectory “unbending” due to the radial component of the magnetic field and the increased multiple scattering). The new measurement significantly reduces the probability of overestimating momentum of soft muons at Level-1 trigger leading to a large reduction of the muon trigger rate (a factor of 6-10) while improving the trigger efficiency due to the increased redundancy of the system. At HL-LHC, in addition to improving Level-1 triggering for the prompt muons by providing well measured muon tracks for matching with the tracks reconstructed by the tracking trigger, the ability to measure directions of muon tracks allows designing muon trigger efficient for signatures with displaced muons predicted in many models beyond Standard Model. The latter is important as the viable HL-LHC tracking trigger designs are inefficient for tracks with the impact parameter exceeding  $d_{xy} \sim 2 - 3$  mm. The ME0 detector will extend the offline coverage of the CMS muon system from  $\eta \sim 2.4$  to close to  $\eta \sim 3$ . In addition, similar to GE1/1 and GE2/1 detectors, the data from the upper part of the ME0 detector will be used for triggering to improve Level-1 momentum measurement and background rejection in the region  $\eta = 2.16 - 2.45$  (GE1/1 extends to  $\eta \sim 2.16$  so the ME0 will close this gap to enable efficient triggering in the entire range covered by CSC). The optimal design requires the data from GEM detectors to be sent to the CSC Optical Trigger Motherboard for joint co-processing of the GEM and CSC trigger data, which requires firmware modifications for the CSC electronics in addition to developing trigger and DAQ systems for the new detectors. As the CMS plans to install the GE1/1 detector during the Long Shutdown 2 (LS2), the scope covered by the document focuses on developing the DAQ/trigger systems for GE2/1 and ME0 and their integration with the CSC electronics and the central CMS DAQ system.

The US role in the GE2/1 and ME0 DAQ/trigger system construction includes development and construction of the Optohybrid (OH) boards that will be located on chambers and which concentrate signals from the readout chips and communicate the data to the CSC trigger electronics and to the dedicated backend DAQ/trigger  $\mu$ TCA systems, links connecting the OH with the CSC and the backend electronics, as well as the backend electronics (crates, processing boards and interfaces). Figure ?? provides an illustration of the layout of the GEM readout system and its main components. The OH board will be connecting to the so-called GEM Electronics Board (GEB), which routes the signals from the chips to the OH board. The GEB board will be built by our international collaborators as well as the VFAT custom front-end chip. VFAT-3 is an



iteration on the existing VFAT-2 chip and it is currently being designed by CERN for GE1/1. VFAT-3 will be reused for GE2/1, while ME0 will use the next generation of VFAT chip.

The GE2/1 trigger/DAQ system will closely follow the design of the GE1/1 system that will be installed in CMS in 2019-2020 during the Long Shutdown 2. The GE1/1 design is approaching final stages, including advanced prototypes of critical electronics components and advanced electronics integration efforts. The main modifications required for the GE2/1 system are related to different mechanical constraints and signal routing for the GE2/1 OH board due to significant geometrical differences of the two detectors. A modest amount of electronics design, prototyping and firmware development will be necessary as part of the RD studies to be performed by qualified engineers.

The ME0 detector will be positioned in a space that will become available at the back of the new Endcap Calorimeter that is being designed now. Much like in the GE1/1 and GE2/1 case, the GEM technology is a uniquely suitable option due to its very compact size, but the high rate capability is another key factor as the radiation levels in this region are substantially higher than the rates that will be seen by the existing CMS muon detectors. The constraints and still existing uncertainties in the final geometric envelope and impact of high radiation rate imply a need for a significant RD effort required to determine the parameters and final technology choices for the ME0 DAQ/trigger system design. The current and planned RD studies include selection of radiation hard components conforming to the highly constrained geometrical envelope as well as understanding of the radiation damage mitigation options, which will have an impact on the choice of the FPGA package and other digital components of the OH board as well as the optimal placement of the board. In addition to electronics prototyping efforts, firmware development will be essential in evaluating and testing the SEU mitigation techniques and establishing compatibility with the existing systems (most notably the CSC ME1/1 OTMB) in terms of latency, algorithmic implementations and data volume. The plan is to develop the first realistic design of the system for the CMS Muon TDR that is currently targeted for release in 2018. The RD studies will need to continue until 2020 to finalize details of the design implementation for the key functional elements and arrive to a first functional prototype. Resources required include fractions of several engineers and computing professionals to work on the electronics design studies, firmware and control and readout software required for performing and evaluating results of the tests. A modest amount of work will also need to be invested in understanding the backend electronics design given the expected evolution in technologies and ensuring compatibility of the backend system with the parameters of the OH board as well as CMS central DAQ and Level-1 trigger systems, which the backend electronics is designed to communicate with.

## 9.4 Performance

## 9.5 Alternatives

Potential alternatives for the CSC off-chamber electronics upgrade include building the backend FED system using already existing in CMS uTCA boards instead of building an TCA-based system. The disadvantage of this approach is in significantly larger amount of engineering effort required and the increased complexity of maintaining multiple electronics formats in the operations stage, including the need for larger amount of spares. For the ODMB board, alternatives that are being considered are mainly focusing on the choice of a specific FPGA. More significant changes to the design would require rebuilding a substantial part of other electronics components (including OTMB, ODMB boards and crates for stations not affected by this



upgrade) that are expected to be preserved under the current plan. Such redesign would significantly increase the cost of the upgrade and would require additional R&D studies. For the GEM detectors, one viable alternative is to design the system to use newer low power GBT chips (LpGBT) instead of GBTX chips. This would require the new LpGBT chips to be available in time for the GE2/1 OH board production and would allow reducing the number of links in the system. Other alternatives include building the backend using existing designs of the processing board in the uTCA format. That would allow an earlier onset of the work to develop final firmware and software for the GEM detectors backend, but is likely to lead to additional costs required to address potential compatibility issues with the yet to be adopted CMS standard platform for Phase-2 upgrades, which is expected to be higher performant version of xTCA. Furthermore, additional maintenance efforts and higher numbers of spares make this option unattractive. For the OH boards, viable alternatives to the default design could come from using a different FPGA package, either among those available now or in the future with the goal of optimizing the cost of the OH board.

## 9.6 Environmental Impact, Health, and Safety

The environmental impact from the electronics upgrades is expected to be extremely small as all part are standard off the shelf items including FPGAs, connectors, clock, and flash ADC, the cables are all non-halogen per CERN regulations. Majority of the activities related to development and testing these electronics items would only imply use of low power (8 Volts or lower) to power the electronics components. As such there is no danger of electrocution. Only select integration activities would require working with the electronics connected to a chamber powered by high voltage. In those cases, standard protocols and safety rules for work with high voltage will be fully enforced and monitored to minimize safety risks. Fire risks are present, as with all electronics systems. As such each board has fuses on the various voltages supplied to the boards. Computers also monitor the current, temperature, and voltage on the boards continuously when powered. Radiation safety considerations affect a small number of specialized activities related to studies of electronics radiation hardness. All of these studies will be carried out either at US facilities or CERN, which have their own very strict requirements and policies related to health and safety. Personnel will follow the regulations and requirements set by the facilities at which such studies are performed, including radiation training, and dosimeter requirements are strictly followed by all of our personnel. There are no specific health risks associated with the use of the electronics items to be developed. Other non-specific risks associated with use of standard lab and office equipment will be addressed by following standard health and safety procedures developed at the participating institutions and facilities where such activities are to take place. In all cases, Project Manager will be directly overseeing all aspects related to health and safety of the personnel involved.

## 9.7 Risk

## 9.8 Quality Assurance and Quality Control

To ensure that the designs being developed meet the corresponding requirements, in each case there is a plan of extensive prototyping and testing, including testing for radiation hardness and specialized torture tests to validate electronics performance and identify rare errors. Once the designs of the electronics components are established, production of the electronics is planned to have a pre-series or pre-production of small quantities of the electronics components that are to undergo extensive testing and validation before full production is approved.

5236 All electronics components will undergo rigorous testing for potential manufacturing defects,  
5237 with the results of all tests being recorded and preserved in a specialized database. A similar  
5238 testing will be repeated upon the delivery of the items to CERN. From the arrival of electronics  
5239 items from the manufacturer, each item will be tracked for its lifetime and the information to be  
5240 stored in a database. All on-chamber electronics boards will undergo a full systems test before  
5241 the chamber is mounted on the detector, again with all test results saved in the database.

## 5242 9.9 Value Engineering

## 5243 9.10 R&D Towards Final Design

DRAFT

DRAFT

## Chapter 10

# Summary - Editor (TBD) 2 pages

## .1 Appendix: Outer tracker common mechanics and services

### .1.1 Mechanics

The mechanical concept of the Phase-2 tracker is the same as that of the current tracker. The tracker is constructed and assembled in surface facilities and installed into CMS in the following order: first the Outer Tracker, then the central beampipe and last the Inner Tracker. Both the central beampipe and the IT are supported from the OT. The IT can be removed, without removing the beampipe, for maintenance or replacement during Extended Technical Stops, whilst the OT is expected to remain in place for the duration of its operational lifetime. Figure 1 shows how the space within the tracking volume is allocated to the different sub-systems.

The structural backbone of the tracker is the Outer Tracker support tube (OTST) located at the outermost radius of the tracker, spanning both ends. The support tube has a wall thickness of 30 mm, consisting of two skins of carbon fibre / epoxy composite and a honeycomb core made

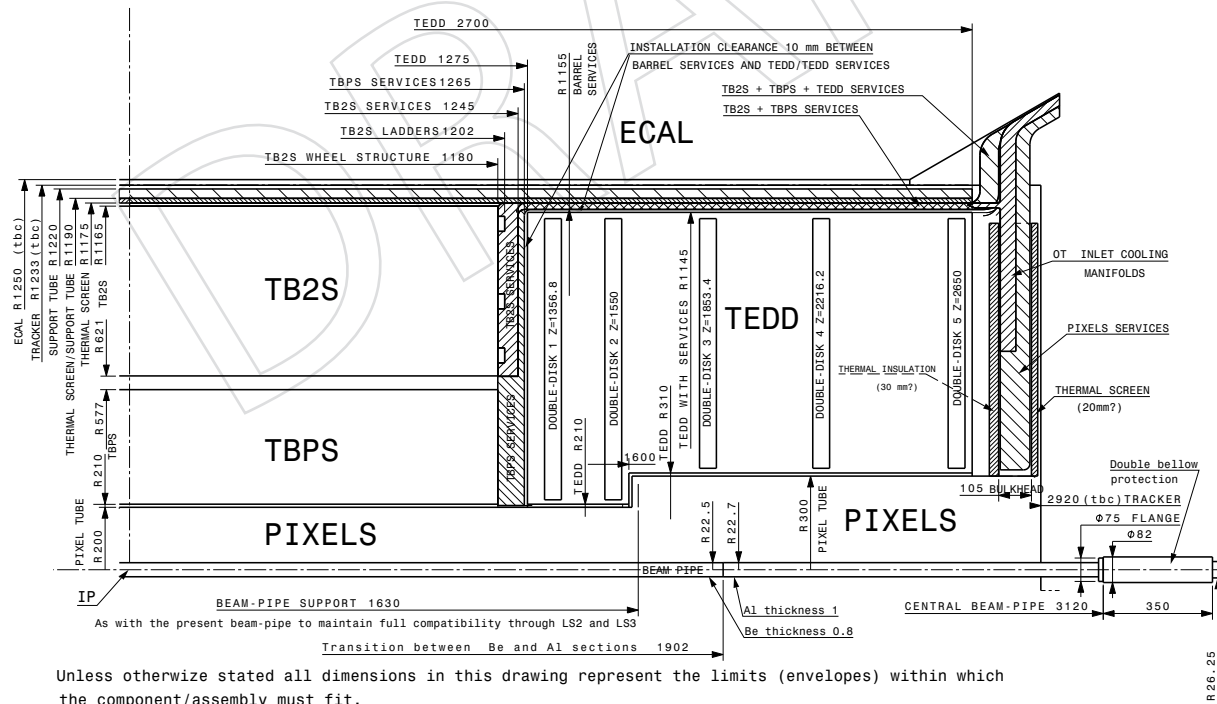


Figure 1: Space allocation in the tracking volume. **FIXME:** The naming in the drawing needs to be updated: change Pixels to Inner Tracker.

of aramid fibres. Metallic attachment inserts are added to the sandwich structure at the ends of the support tube. Two carbon fibre rails at 3 and 9 o'clock positions inside the support tube guide and hold the TB2S and the two TEDD units. Bulkheads, made in carbon fibre composite, close the ends of the OT volume and provide supports for the cooling inlet manifolds, the Inner Tracker support tube (ITST), the central beampipe and the IT services. The ITST provides installation and support rails at its 6 and 12 o'clock positions for the IT and its services. The IT sub-sections have small high-precision support wheels for the installation along the ITST rails. The ITST is made of high-stiffness carbon fibre composite and it is sub-divided into five sections for assembly reasons. However, the fully installed ITST is a continuous shell structure, forming a barrier between the Inner Tracker and Outer Tracker volumes.

The tracker is supported from the CMS barrel hadron calorimeter (HCAL) via four metallic brackets. The tracker has no mechanical connections to its closest neighbour, the barrel electromagnetic calorimeter (ECAL). During installation into CMS the tracker is aligned with respect to CMS and LHC survey references. Once the Outer Tracker cables and pipes are connected, there will be no possibility to further adjust the positions of the OT or of the IT support tube, which moves with the OT. Later corrections can be done, within a range of about  $\pm 3$  mm, by adjusting the positions of the central beampipe and of the Inner Tracker sub-sections. For continuity reasons adjusting the central beampipe requires adjusting also the positions of the endcap beampipe sections in the CMS cavern. The positions of the IT sub-sections can be adjusted by moving their support wheel positions before installation.

For improved radiation resistance the tracker will be operated and maintained at  $-20^\circ\text{C}$  or lower using two-phase  $\text{CO}_2$  cooling. All cold volumes are sealed from surrounding air moisture and continuously flushed with dry air or nitrogen. The most critical sealing is at the ends of the tracker, where the cables and pipes are routed outwards. Silicone potting and foam materials are used to seal the service channels. Insulated thermal screen panels with aluminized vapour barrier foils and aluminium film sealants are used to cover the bulkheads and service channels that need to remain openable.

## .1.2 Services

All the existing tracker services (low and high voltage power cables, optical fibres, cooling pipes, dry gas pipes, sniffing pipes, humidity and temperature measurement cables) located in the CMS detector volume will be removed during Long Shutdown 3 and be replaced by new ones. Before installing the tracker the new services will be laid out on the CMS central wheel (YB0). These long cables and pipes shall reach the patch panels located on the inner wall of the magnet cryostat (referred to as patch panel 1, or PP1). The power cables and optical fibre bundles of the Outer Tracker have pigtails reaching the PP1s. The only services of the OT having connectors at the bulkheads at the trackers' end are the cooling and dry gas pipes.

The Inner Tracker has connectors for all its services at the bulkheads, therefore facilitating the installation and removal of the IT at later stages. To further facilitate access, the IT services will be routed to the service channels closest to the horizontal plane and remain therefore more easily reachable from the IT installation platforms.

The services between the tracker bulkheads and the PP1s are laid into the existing cable channels, shown in Fig. 2, that run on top of the ends of the barrel HCAL and barrel ECAL. The types and quantities of services that are planned to be installed into these cable channels are listed in Tab. 1.



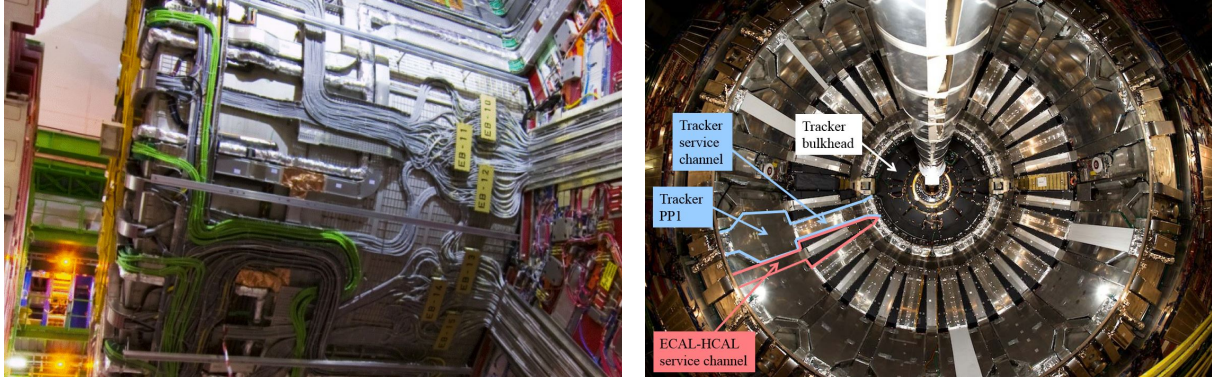


Figure 2: Left: photo of the currently installed services on the CMS central wheel, YB0. Right: photo high-lighting in blue one of the tracker service channels and PP1s, and in pink one of the neighbouring ECAL+HCAL service channels.

Table 1: Quantities and estimated dimensions of the Inner Tracker and Outer Tracker services running from the bulkheads to the PP1s, for both ends of the tracker. **FIXME:** Check and complete the table.

Service type	Quantity, IT	Quantity, OT	Dimension [mm]
Power cables	?	?	Ø 13.4
Multifibre optical cables	?	?	?
Inlet cooling pipes	?	46	?
Outlet cooling pipes	?	46	?
Dry gas injection pipes	8	24	Ø 10
Sniffing pipes	8	24	Ø 6
Cables for measurement of relative humidity and temperature	?	?	?

### .1.3 Tracker assembly sequence

The tracker assembly and commissioning work is done as much as possible in advance, before lowering and insertion of the components into CMS. Outer Tracker sub-detectors are brought to the Tracker Integration Facility (TIF) in building 186 on the CERN Meyrin site, where they are installed into the Outer Tracker support tube. Final service pigtailed (i.e. power and optical cables) as well as cooling supply and return connections are connected to supply systems in the TIF allowing testing of sub-sections of the tracker, including functional tests with cosmic rays.

The completed and tested Outer Tracker is then transported to LHC point 5 (P5), lowered via the main shaft into the experimental cavern and installed into the CMS detector. The OT is aligned with respect to the survey reference system in the CMS cavern. Targets supported by the Inner Tracker guide rails are used as alignment references, with the aim of centering the IT support rails as symmetrically as possible around the expected LHC beam position. As the result of the alignment the OT is then firmly attached to the barrel HCAL via the four support brackets. The OT installation is completed by connecting its cables and pipes to the supply lines installed on YB0, accompanied by functional testing of each service line. This is followed by a checkout and commissioning of the full OT, including test runs with cosmic rays.

The central beampipe will be installed, connected to the endcap beampipes and baked out before installation of the Inner Tracker. The IT is installed in eight sub-sections, each one of them

having been tested before in the surface laboratories, either in the TIF or at the CMS experimental site. Alignment of the beampipe and the IT is done again using the survey network in the CMS cavern, and targeting a precise positioning around the expected LHC beam position. The objective for the final positioning precision of the IT with respect to the beam is  $\pm 1$  mm.

#### .1.4 Cooling system

The total power dissipated in the tracking volume and due to heat leaks from the surroundings are expected to amount to about 100 kW for the Outer Tracker and to about 50 kW for the Inner Tracker, including losses on cables inside the tracking volume and allowing for some margin. The cooling system must remove this heat load and maintain the silicon sensors at a temperature of  $-20^{\circ}\text{C}$  or lower. Two-phase  $\text{CO}_2$  is chosen as the coolant. The cooling system will be designed for a nominal coolant operating temperature of  $-35^{\circ}\text{C}$ , resulting in a coolant temperature of about  $-33^{\circ}\text{C}$  at the location of the first silicon module.

Like in the CMS pixel Phase-1 Upgrade the cooling of the Phase-2 tracker will be based on evaporative  $\text{CO}_2$  in a liquid pumped cycle. A common system is designed to serve both the Outer Tracker and the Inner Tracker. The fluid properties of  $\text{CO}_2$  make it an ideal fluid for cooling of tracking detectors, in particular due to the ability of using smaller diameter, lower-mass tubing than is required with conventional refrigerants or liquid cooling applications [? ]. The first principle reason for this is that  $\text{CO}_2$  evaporates at higher pressures than regular refrigerants. High pressure keeps the vapours compressed and therefore the circuit volume low. The boiling temperature is a function of the pressure and, as the dynamic pressure drops along the cooling pipes, can be kept small compared to the high absolute pressure of the fluid. This results in a minor change in the evaporation temperature along a cooling pipe. An additional factor for this is that  $\text{CO}_2$  has a high latent heat, translating into a smaller needed flow with respect to the one required with other refrigerants. As the viscosity of  $\text{CO}_2$  is low it allows use of small diameter pipes with higher flow speeds which increase the heat transfer coefficient from the pipe wall to the fluid.  $\text{CO}_2$  is radiation hard, it is cheap and environmentally friendly.

The drawback of  $\text{CO}_2$  is the high pressure and the associated safety aspects; nonetheless due to the small acceptable pipe diameters the overall stored energy, which is the product of volume and pressure, is similar to the one of low-pressure systems with larger volumes.

All current  $\text{CO}_2$  cooling systems at CERN are based on the 2-Phase Accumulator Controlled Loop concept (2PACL). In this system, in comparison to usual two-phase compressor systems, only part of the circulating fluid is evaporated and the return flow from the detector is a mixture of liquid and evaporated vapour. This cooling system concept has been successfully employed in the LHCb VELO since 2008, and is since 2017 in use in the Phase-1 CMS pixel detector. This design provides a stable cooling temperature in the detector over a large temperature range from room temperature down to  $-35^{\circ}\text{C}$ . The cooling temperature is nearly independent of the heat load. The system design is such that all active hardware is away from the detector in accessible areas. The hardware amount inside the detector's acceptance is minimal and consists mainly of small diameter tubing. This all makes the  $\text{CO}_2$  2PACL approach well suited for cooling of tracking detectors.

The  $\text{CO}_2$  cooling plants, located in the service cavern (USC), provide a flow of  $\text{CO}_2$  in liquid phase. The flow is distributed through manifolds located in the experimental cavern (UXC) to 46 Outer Tracker and NN **FIXME: Add the number.** Inner Tracker cooling segments. Each of the cooling segments is then further split into parallel detector cooling loops at the entry to the tracker volume. All detector cooling pipes are preceded by capillaries, which create the needed pressure drop to reach fluid saturation conditions and guarantee a uniform distribution. To

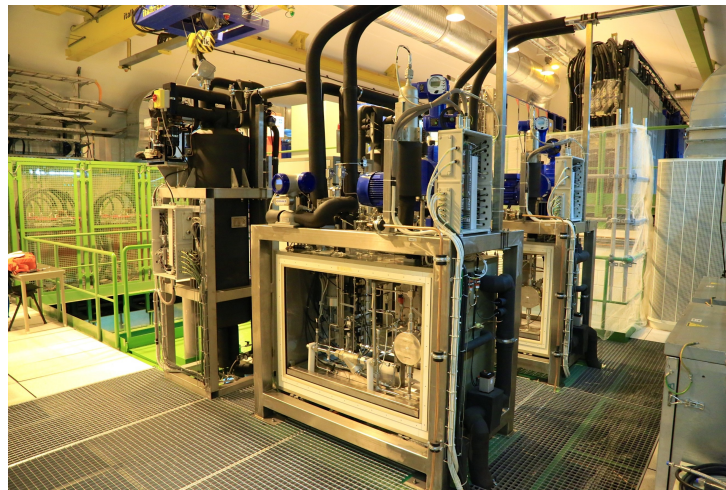


Figure 3: The cooling plants of the pixel Phase-1 CO<sub>2</sub> cooling system at CERN. A similar system, but larger in scale, will be used for the Phase-2 tracker.

ensure onset of evaporation, small preheaters with an adjustable power corresponding to one to two detector modules (5-10 W) are installed upstream of the first 2S, PS or pixel module in each cooling loop. Every preheater is powered through the same power cable as the neighbouring modules and is thus switched on and off simultaneously with the detector powering.

To avoid risks of dryout and loss of cooling, only about 50% of the fluid is evaporated. The return flow is thus a mixture of liquid and gaseous CO<sub>2</sub>. The two CO<sub>2</sub> cooling plants currently installed for the Phase-1 pixel detector have each a cooling capacity of 15 kW (Fig. 3). The plan for the Phase-2 Upgrade is to construct cooling plants of about 50 kW maximum cooling capacity. The basic concept is to have five identical cooling plants, all working at about 75% of their capacity during normal operation. In case of a failure or during maintenance periods the still operational plants can be operated at full power and take over the load.

Thermal insulation panels with heating foils on their surface are used to maintain the periphery of the tracker close to the temperature of the surrounding detectors (expected to be +18°C or slightly lower), and always above the dew point of the surrounding air. When CMS is open the tracker is exposed to the cavern air the dew point of which can be up to +13°C. No active cooling panels are placed inside the insulation layer. The detector cooling plants, or at least a sufficient subset of them, will always be in operation to keep the tracker volume at low temperature, as required to avoid incurring additional damage from radiation effects as a result of reverse annealing. Diesel-powered electrical supplies are used to ensure continuous operation even during power cuts for two plants out of five and for their primary system. Particular care will be devoted to the engineering of the bulkheads and the feedthroughs for the service lines, in order to ensure a proper sealing of the cold volume. The tracker volume will be continuously flushed with nitrogen, or dry air for safety reasons when CMS is open, as it is done for the present tracker.

### .1.5 Central beampipe

A new central beampipe is needed in order to gain space for the new, enlarged Inner Tracker with high pseudorapidity coverage. Figure 1 shows the main dimensions of the new beam pipe. While the present beampipe has conical end sections, the new beampipe will be fully cylindrical. The diameter of the pipe will be 45 mm, which is the same as in the cylindrical mid section of the present beampipe. The new beampipe will be either fully in 0.8 mm thick

5397 beryllium, or will be of beryllium in the central section and of aluminium (Al2219 alloy) in the  
5398 end sections. The end flanges will be in Al2219 with local stainless steel reinforcements in the  
5399 vacuum sealing.

5400 The new beampipe is planned to be installed during Long Shutdown 2 (2019-2020) to match  
5401 with the replacement of the CMS endcap beampipes. During LS3 these beampipes will be  
5402 removed, stored, and reinstalled again at the end of the CMS assembly sequence.

DRAFT



## Appendix A

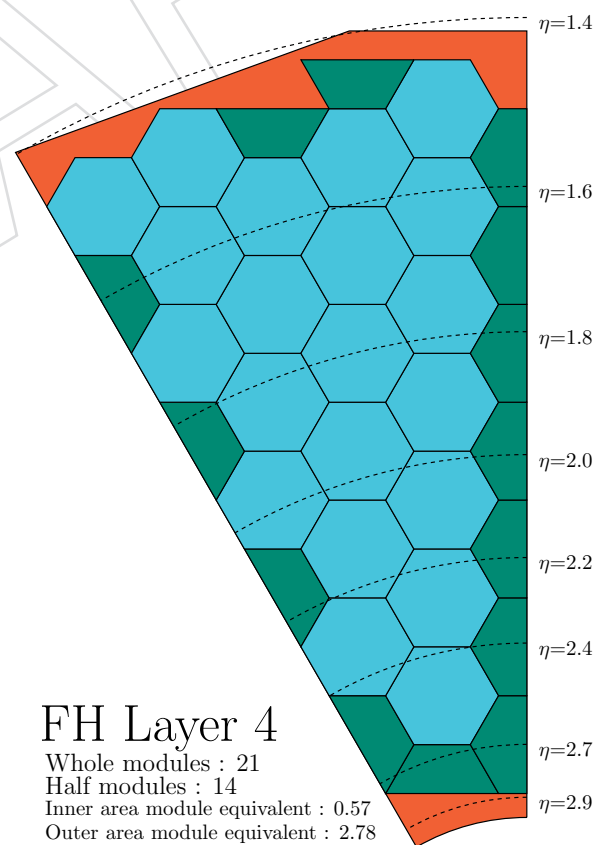
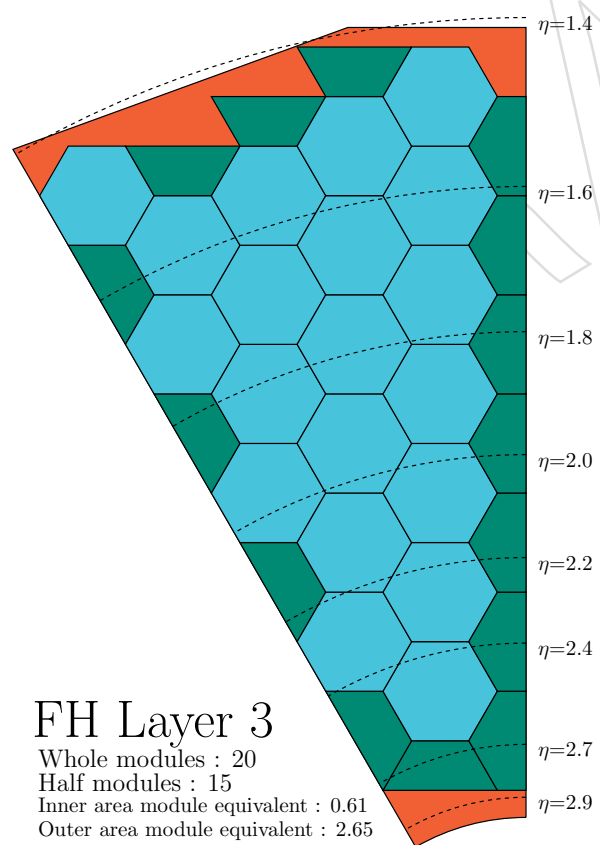
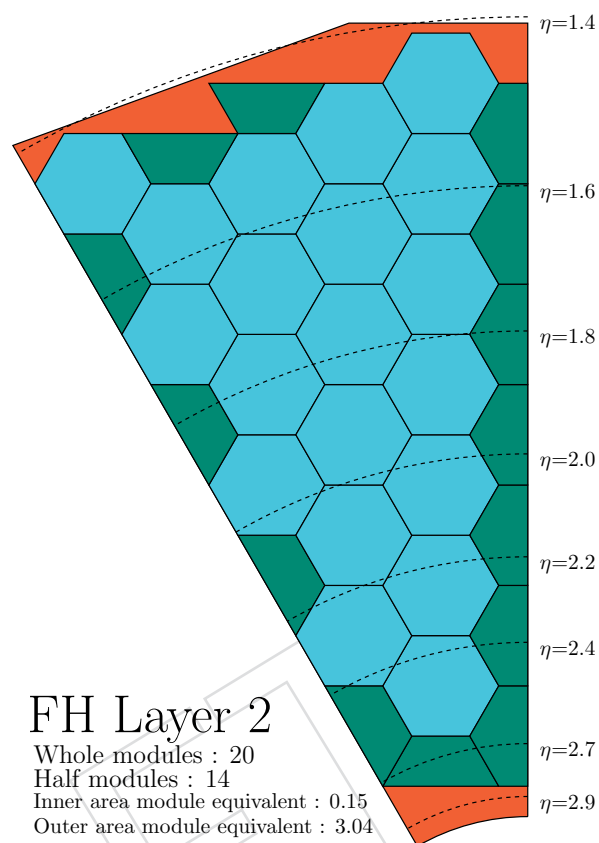
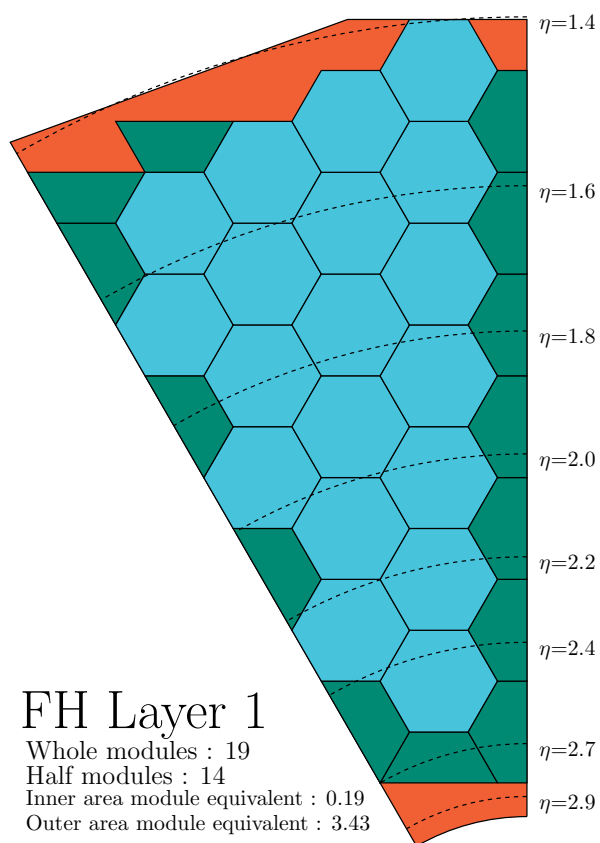
# Endcap Calorimeter Supplementary Material

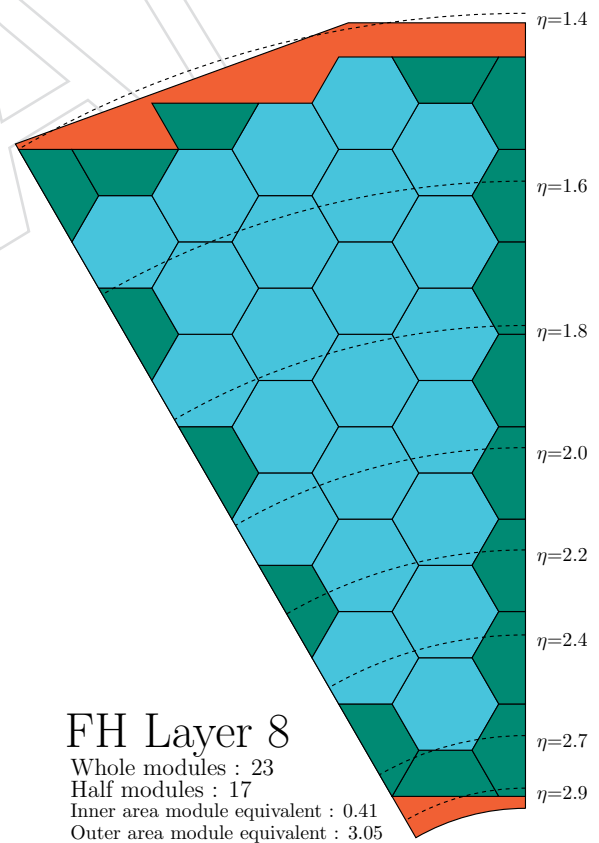
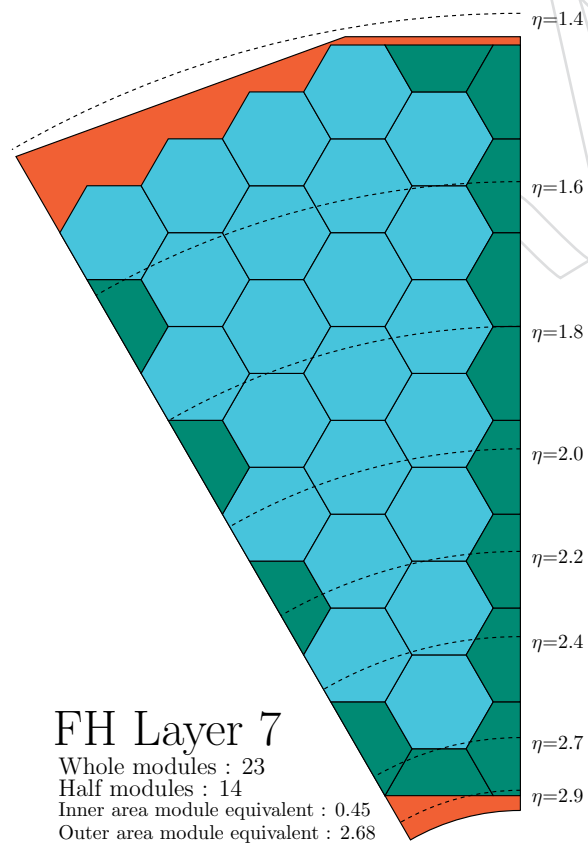
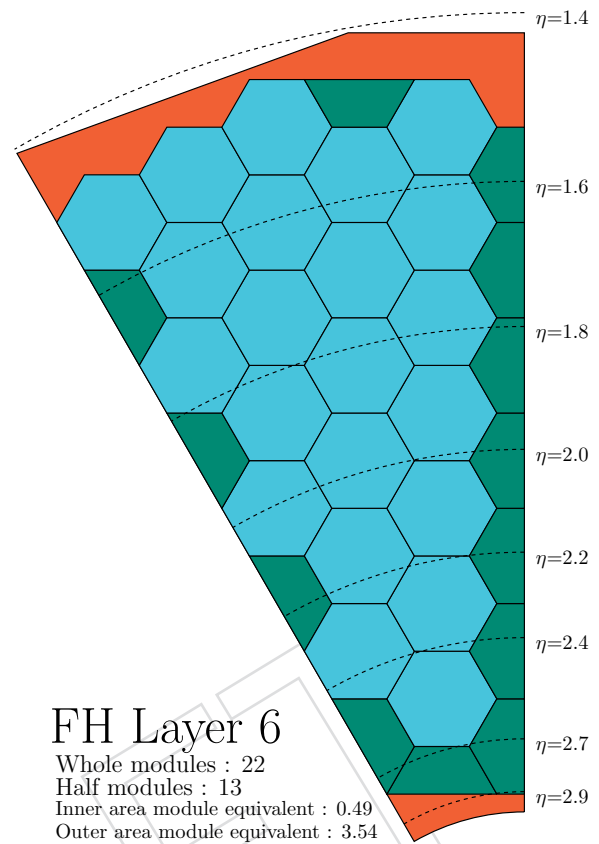
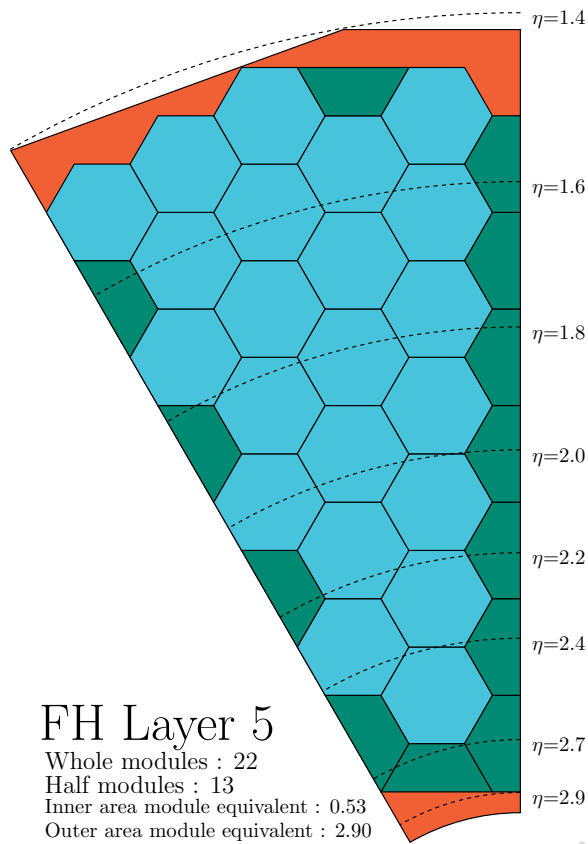
## A.1 Cassette layout information

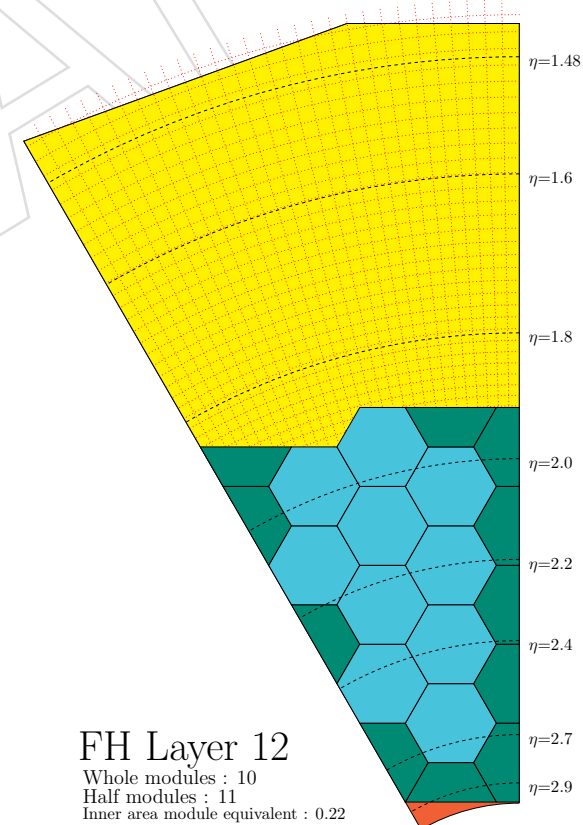
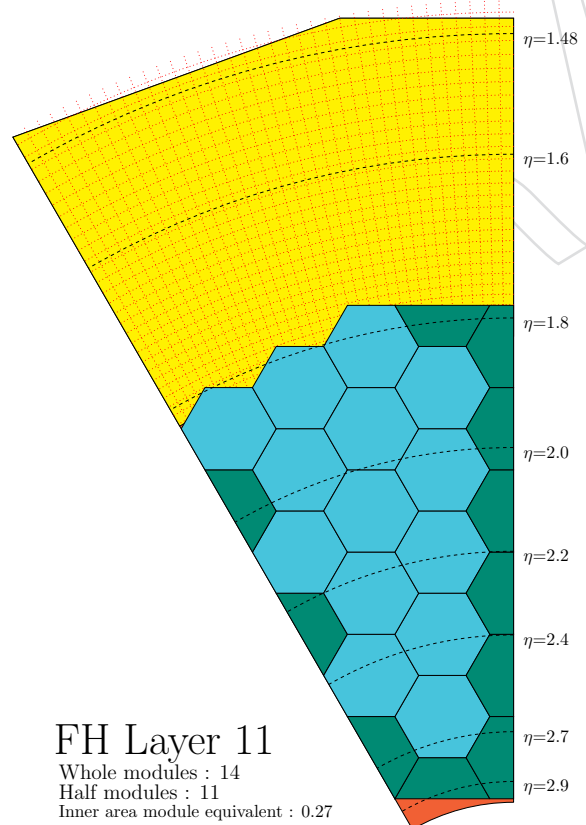
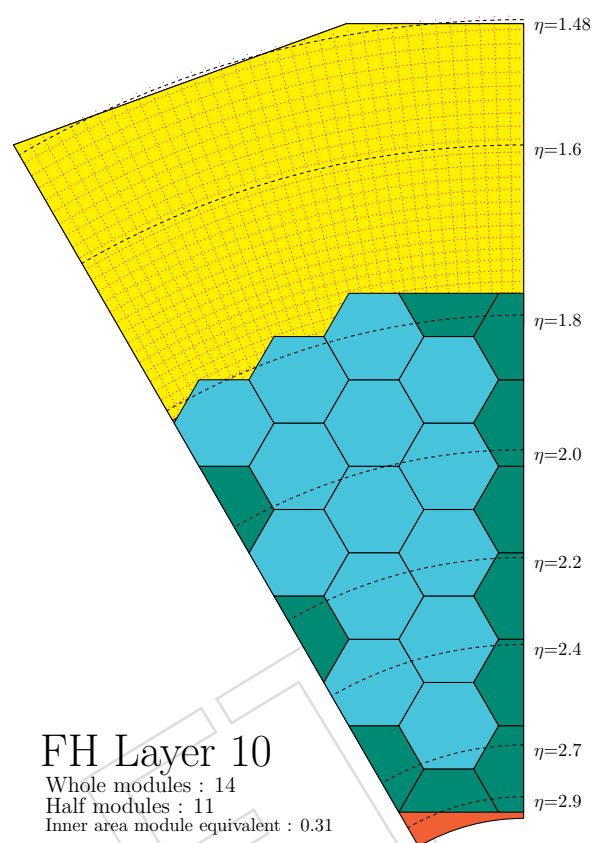
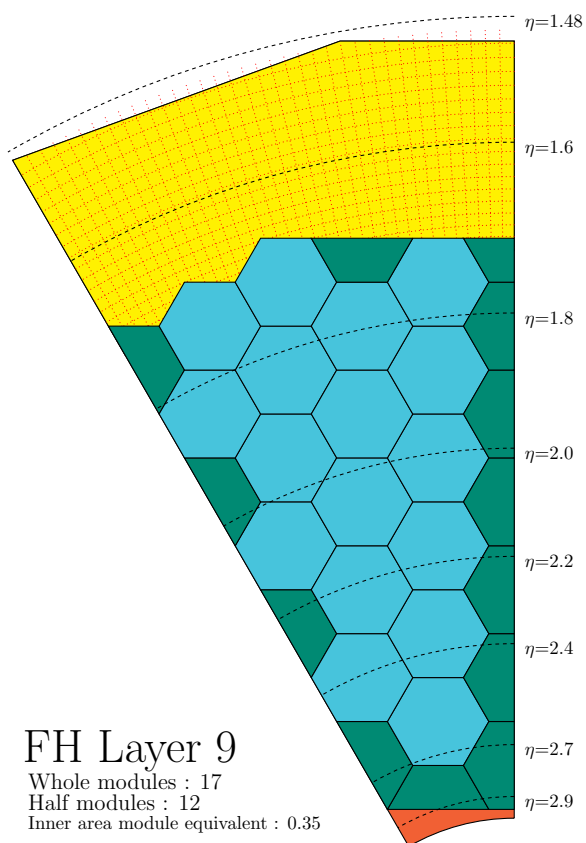
Table A.1: Per-layer counts of modules of various types in the endcap calorimeter

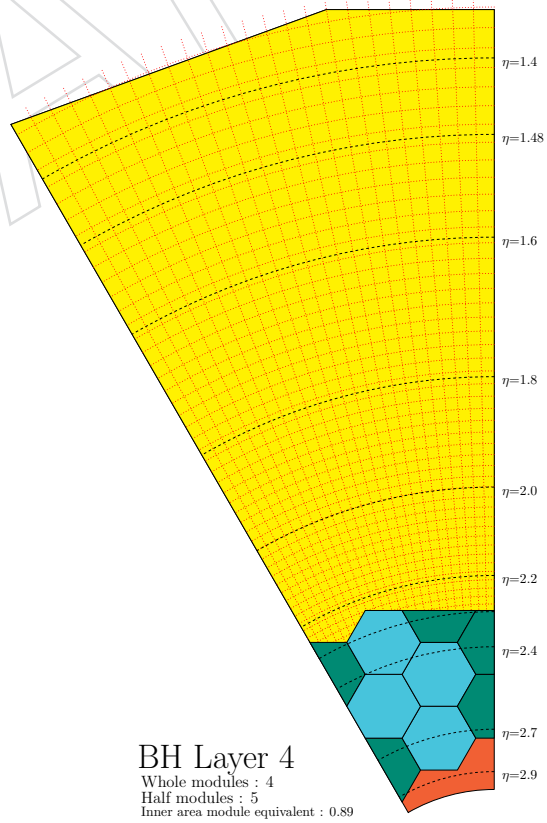
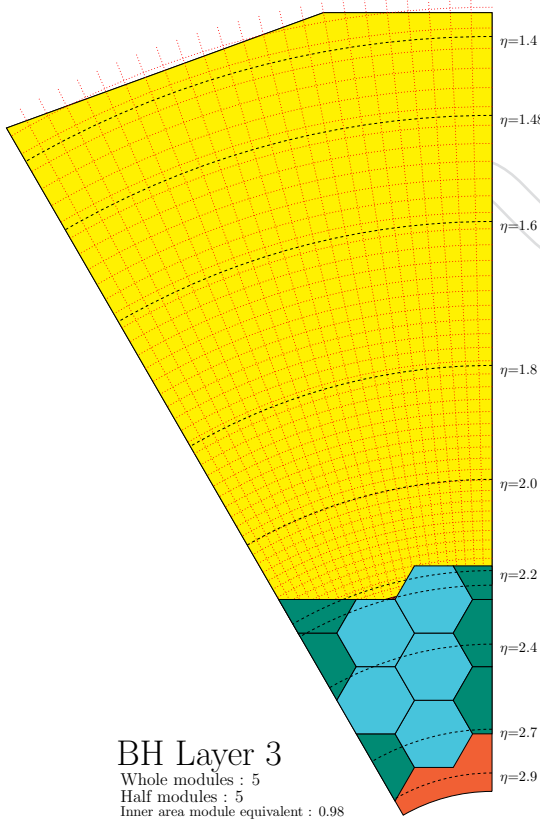
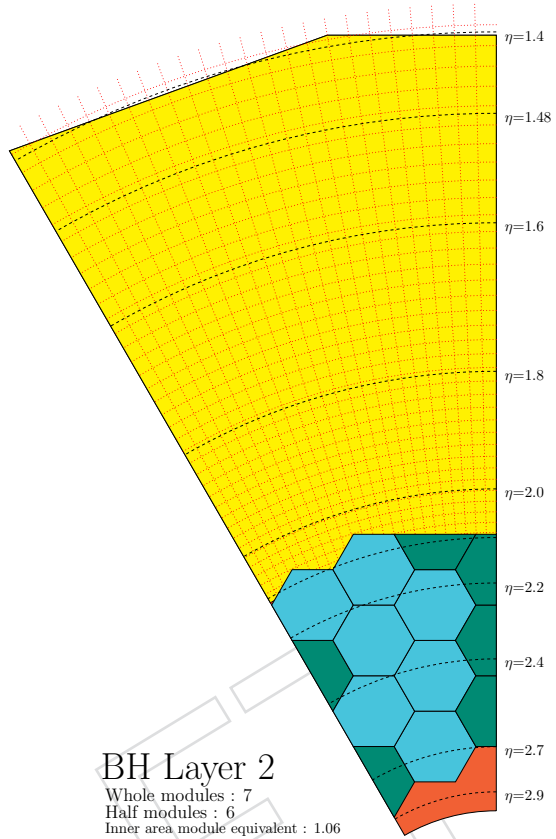
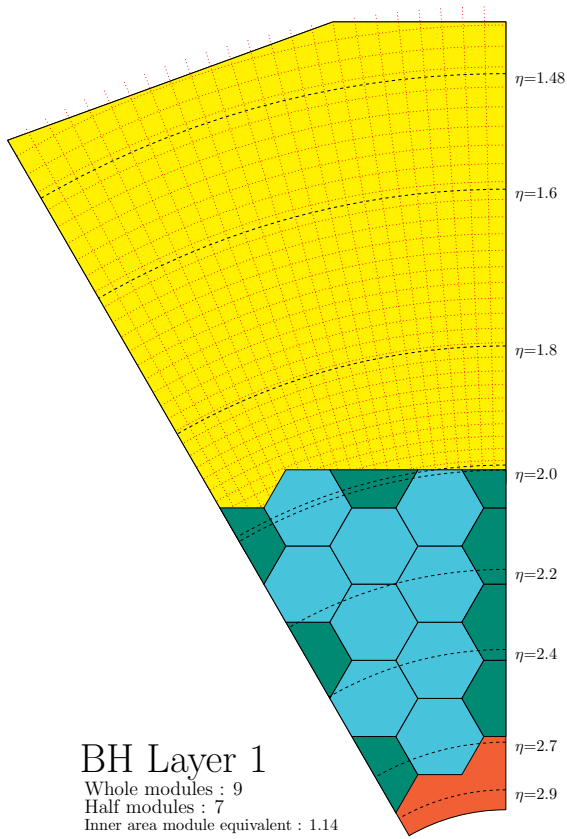
Layer	Whole modules	Half Modules	Partial module area	Total full-size equivalent
FH1	456	336	86.88	710.88
FH2	480	336	76.56	724.56
FH3	480	360	78.24	738.24
FH4	504	336	80.4	752.4
FH5	528	312	82.32	766.32
FH6	528	312	96.72	780.72
FH7	552	336	75.12	795.12
FH8	552	408	83.04	839.04
FH9	408	288	14.4	566.4
FH10	336	264	7.44	475.44
FH11	336	264	6.48	474.48
FH12	240	264	5.28	377.28
BH1	216	168	33.36	333.36
BH2	168	144	25.44	265.44
BH3	120	120	29.52	209.52
BH4	96	120	21.36	177.36
BH5	96	120	19.44	175.44
BH6	96	120	17.28	173.28
BH7	96	120	15.12	171.12
BH8	96	96	24.96	168.96
BH9	96	96	22.56	166.56
BH10	96	96	20.4	164.4
BH11	96	120	36	192
BH12	96	120	33.84	189.84
Total	6768	5256	992.16	10388.16

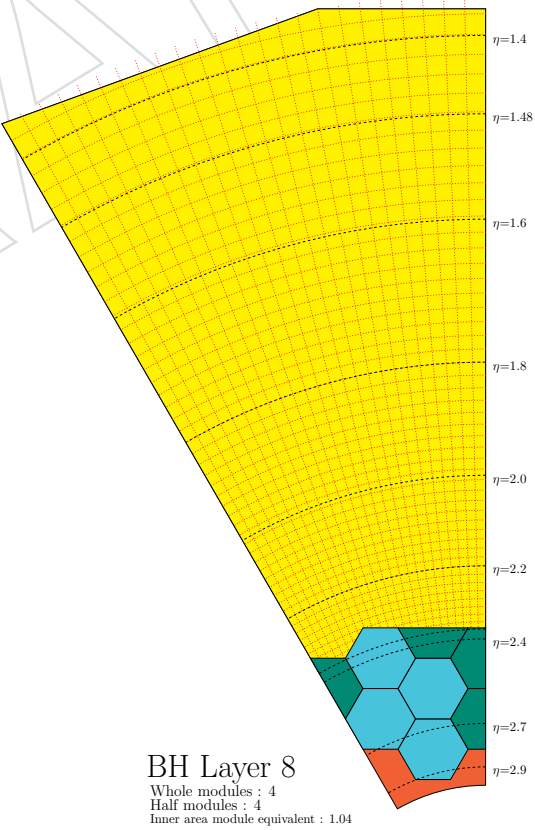
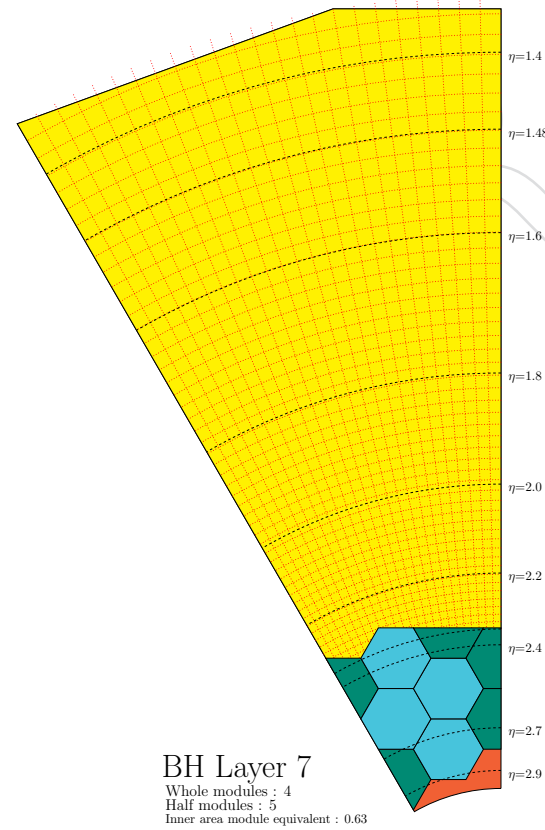
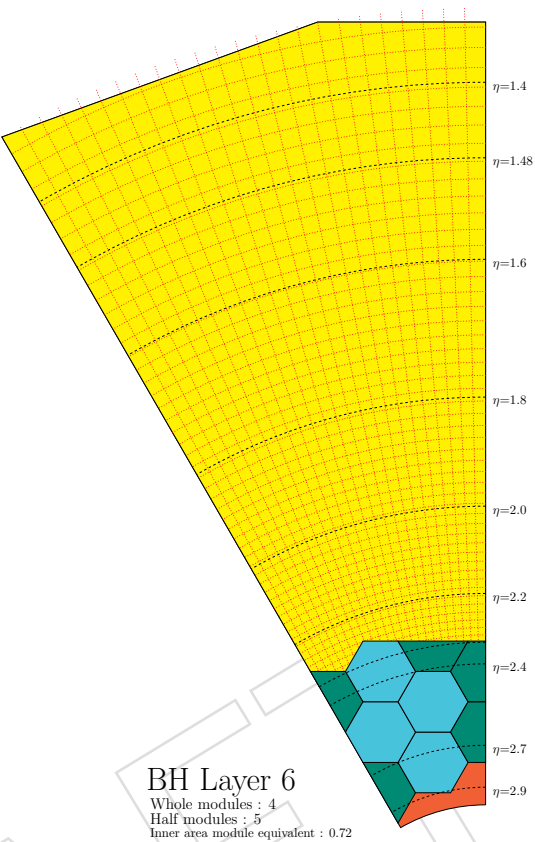
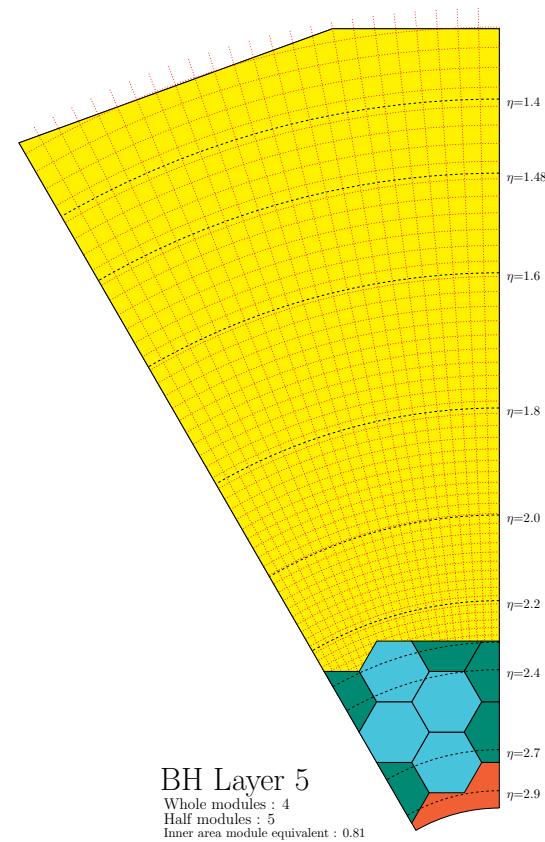






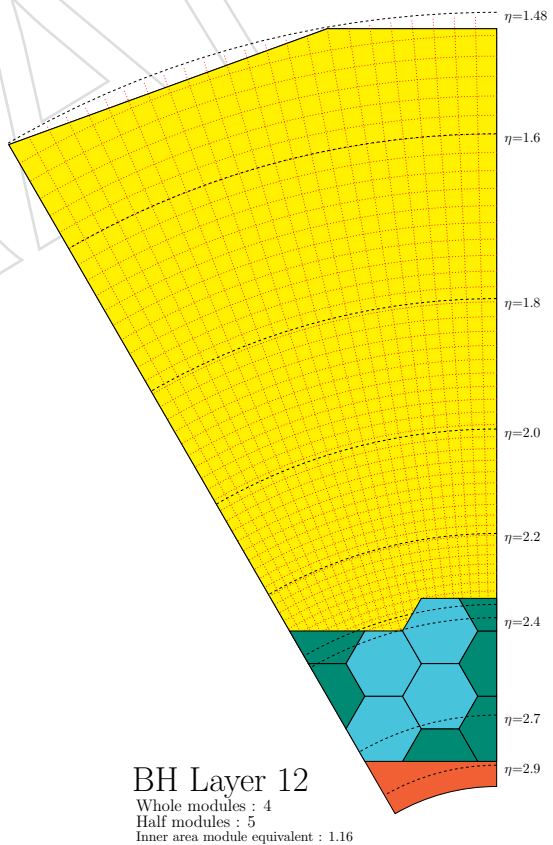
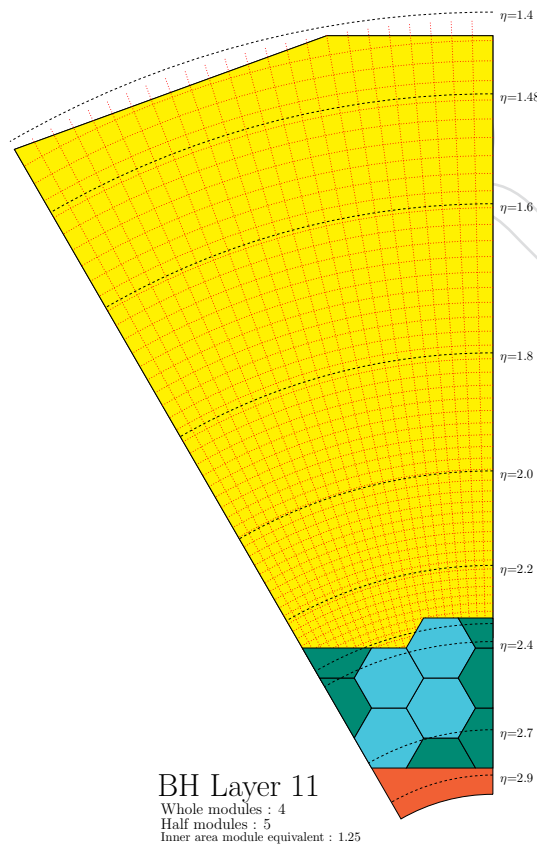
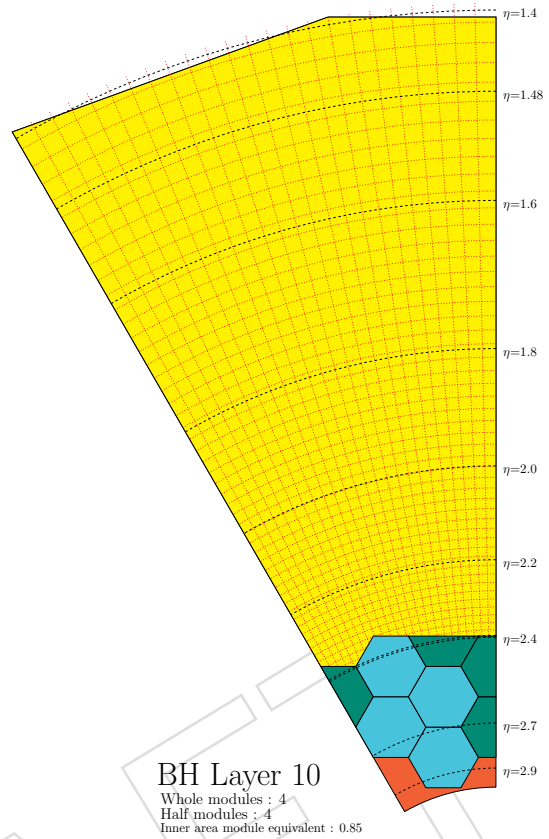
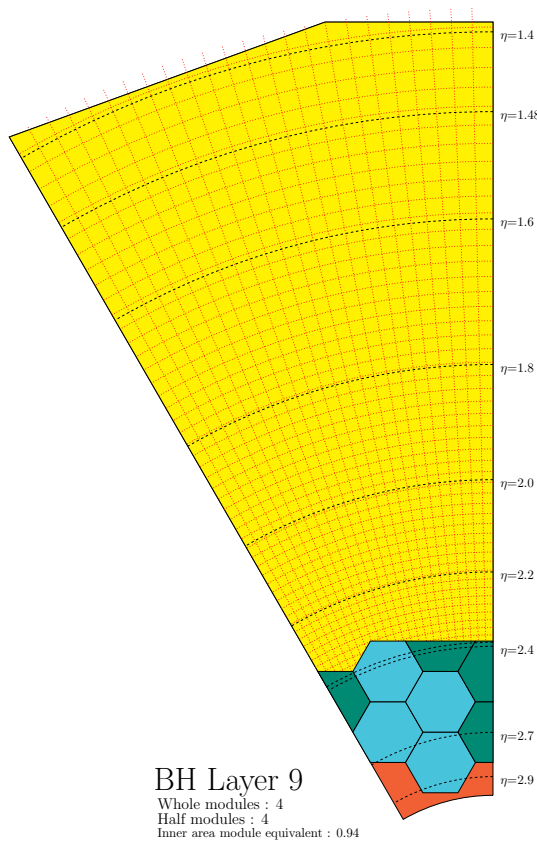






5410





DRAFT

## References

- 
- [1] ATLAS Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”, *Phys. Lett. B* **716** (2012) 1, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.
- [2] CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”, *Phys. Lett. B* **716** (2012) 30, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.
- [3] CMS Collaboration, “Precise determination of the mass of the Higgs boson and studies of the compatibility of its couplings with the standard model”, Technical Report CMS-PAS-HIG-14-009, CERN, Geneva, 2014.
- [4] CMS Collaboration, “Projected Performance of an Upgraded CMS Detector at the LHC and HL-LHC: Contribution to the Snowmass Process”, arXiv:1307.7135.
- [5] S. Malik et al., “Interplay and Characterization of Dark Matter Searches at Colliders and in Direct Detection Experiments”, arXiv:1409.4075.
- [6] I. Azhgirey and V. Talanov, “Proc. of XVIII Workshop on the Charged Particle Accelerators”, Technical Report CMS-NOTE-2005-010, 2000.
- [7] G. Battistoni et al., “The FLUKA code: Description and benchmarking”, *AIP Conf. Proc.* **896** (2007), no. SLAC-REPRINT-2007-184, 31.
- [8] A. Ferrari, P. R. Sala, A. Fassò, and J. Ranft, “FLUKA: A multi-particle transport code (program version 2005)”. CERN, Geneva, 2005.
- [9] S. Roesler, R. Engle, and J. Ranft, “The Monte Carlo Event Generator DMPJET III”, Technical Report SLAC-PUB-8740, 2000.
- [10] T. Pierog et al., “EPOS LHC : test of collective hadronization with LHC data”, Technical Report arXiv:1306.0121, Jun, 2013. Comments: 16 pages, 34 figures.
- [11] I. Bergstrom and H. Vincke, “Residual dose rates in CMS during long shutdowns following the LHC operation until LS5”, Technical Report CERN-RP-2014-052-REPORTS-TN, CERN, 2014.
- [12] A. Dabrowski et al., “Development of the CMS nominal FLUKA Model; versions 2.0.0.0 and 3.0.0.0”, Technical Report AAA-AA-AA-0002, CERN, 2014.
- [13] CMS Collaboration, “The CMS experiment at the CERN LHC”, *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

- [14] CMS and ATLAS Collaborations, “Expected pileup values at the HL-LHC”, Technical Report CMS-NOTE-2013-003, 2013.
- [15] C. Adloff et al., “Response of the CALICE Si-W electromagnetic calorimeter physics prototype to electrons”, *Nucl.Instrum.Meth.* **A608** (2009) 372, doi:10.1016/j.nima.2009.07.026.
- [16] J. de Favereau et al., “DELPHES 3, A modular framework for fast simulation of a generic collider experiment”, arXiv:1307.6346.
- [17] CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”, *Phys. Lett.* **716** (2012) 30, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.
- [18] CMS Collaboration, “Observation of a new boson with mass near 125 GeV in pp collisions at  $\sqrt{s} = 7$  and 8 TeV”, *JHEP* **06** (2013) 081, arXiv:1303.4571.
- [19] ATLAS Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”, *Phys. Lett. B* **716** (2012) 1, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.
- [20] R. S. Gupta, H. Rzehak, and J. D. Wells, “How well do we need to measure Higgs boson couplings?”, *Phys. Rev. D* **86** (2012) 095001, doi:10.1103/PhysRevD.86.095001, arXiv:1206.3560.
- [21] TLEP Design Study Working Group Collaboration, “First Look at the Physics Case of TLEP”, *JHEP* **1401** (2014) 164, doi:10.1007/JHEP01(2014)164, arXiv:1308.6176.
- [22] CMS Collaboration, “Evidence for the direct decay of the 125 GeV Higgs boson to fermions”, *Nature Phys.* **10** (2014) 557, doi:10.1038/nphys3005, arXiv:1401.6527.
- [23] CMS Collaboration, “Technical Proposal for the Upgrade of the CMS Detector through 2020”, Technical Report CERN-LHCC-2011-006, CMS-UG-TP-1, LHCC-P-004, 2011.
- [24] CMS Collaboration, “Updated measurements of Higgs boson production in the diphoton decay channel at  $\sqrt{s} = 13$  TeV in pp collisions at CMS.”, Technical Report CMS-PAS-HIG-16-020, CERN, Geneva, 2016.
- [25] CMS Collaboration, “Measurements of properties of the Higgs boson and search for an additional resonance in the four-lepton final state at  $\sqrt{s} = 13$  TeV”, Technical Report CMS-PAS-HIG-16-033, CERN, Geneva, 2016.
- [26] J. Baglio et al., “The measurement of the Higgs self-coupling at the LHC: theoretical status”, *JHEP* **1304** (2013) 151, doi:10.1007/JHEP04(2013)151, arXiv:1212.5581.
- [27] J. Grigo, K. Melnikov, and M. Steinhauser, “Virtual corrections to Higgs boson pair production in the large top quark mass limit”, *Nucl. Phys. B* **888** (2014) 17, doi:10.1016/j.nuclphysb.2014.09.003, arXiv:1408.2422.
- [28] CMS Collaboration, “Study of the Discovery Reach in Searches for Supersymmetry at CMS with 3000 fb<sup>-1</sup>”, CMS Physics Analysis Summary CMS-PAS-FTR-13-014, 2013.

- [29] CMS Collaboration, “Supersymmetry discovery potential in future LHC and HL-LHC running with the CMS detector”, CMS Physics Analysis Summary CMS-PAS-SUS-14-012, 2014.
- [30] M. Papucci, J. T. Ruderman, and A. Weiler, “Natural SUSY endures”, *JHEP* **1209** (2012) 035, doi:10.1007/JHEP09(2012)035, arXiv:1110.6926.
- [31] J. Alwall, P. Schuster, and N. Toro, “Simplified models for a first characterization of new physics at the LHC”, *Phys. Rev. D* **79** (2009) 075020, doi:10.1103/PhysRevD.79.075020, arXiv:0810.3921.
- [32] D. Alves et al., “Simplified models for LHC new physics searches”, *J. Phys. G* **39** (2012) 105005, doi:10.1088/0954-3899/39/10/105005, arXiv:1105.2838.
- [33] CMS Collaboration, “CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter”, Technical Report CERN-LHCC-2012-015, CMS-TDR-010, CERN, 2012.
- [34] CMS Collaboration, “CMS Technical Design Report for the Pixel Detector Upgrade”, CMS Technical Design Report CERN-LHCC-2012-016, CMS-TDR-11, CERN, 2012.
- [35] CMS Collaboration, “CMS Technical Design Report for the Level-1 Trigger Upgrade”, CMS Technical Design Report CERN-LHCC-2013-011, CMS-TDR-12, 2013.
- [36] DELPHES 3 Collaboration, “DELPHES 3, A modular framework for fast simulation of a generic collider experiment”, *JHEP* **1402** (2014) 057, doi:10.1007/JHEP02(2014)057, arXiv:1307.6346.
- [37] T. Sjöstrand, S. Mrenna, and P. Skands, “PYTHIA 6.4 physics and manual”, *JHEP* **05** (2006) 026, doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.
- [38] J. Alwall et al., “MadGraph 5 : Going Beyond”, *JHEP* **1106** (2011) 128, doi:10.1007/JHEP06(2011)128, arXiv:1106.0522.
- [39] J. Anderson et al., “Snowmass Energy Frontier Simulations”, arXiv:1309.1057.
- [40] A. Avetisyan et al., “Methods and Results for Standard Model Event Generation at  $\sqrt{s} = 14$  TeV, 33 TeV and 100 TeV Proton Colliders (A Snowmass Whitepaper)”, arXiv:1308.1636.
- [41] A. Avetisyan et al., “Snowmass Energy Frontier Simulations using the Open Science Grid (A Snowmass 2013 whitepaper)”, arXiv:1308.0843.
- [42] O. Buchmueller et al., “Likelihood functions for supersymmetric observables in frequentist analyses of the CMSSM and NUHM1”, *Eur. Phys. J. C* **64** (2009) 391, doi:10.1140/epjc/s10052-009-1159-z, arXiv:0907.5568.
- [43] J. R. Ellis, K. A. Olive, and Y. Santoso, “Calculations of neutralino stop coannihilation in the CMSSM”, *Astropart. Phys.* **18** (2003) 395, doi:10.1016/S0927-6505(02)00151-2, arXiv:hep-ph/0112113.
- [44] J. Ellis, K. A. Olive, and J. Zheng, “The Extent of the Stop Coannihilation Strip”, arXiv:1404.5571.



- [45] CMS Collaboration, “Search for top squarks decaying to a charm quark and a neutralino in events with a jet and missing transverse momentum”, CMS Physics Analysis Summary CMS-PAS-SUS-13-009, 2013.
- [46] A. Barr, C. Lester, and P. Stephens, “A variable for measuring masses at hadron colliders when missing energy is expected;  $m_{T2}$ : the truth behind the glamour”, *J. Phys. G* **29** (2003) 2343, doi:10.1088/0954-3899/29/10/304, arXiv:arXiv:hep-ph/0304226.
- [47] M. Burns, K. Kong, K. T. Matchev, and M. Park, “Using subsystem  $M_{T2}$  for complete mass determinations in decay chains with missing energy at hadron colliders”, *JHEP* **0903** (2009) 143, doi:10.1088/1126-6708/2009/03/143, arXiv:arXiv:0810.5576.
- [48] D. Tovey, “On measuring the masses of pair-produced semi-invisibly decaying particles at hadron colliders”, *JHEP* **0804** (2008) 034, doi:10.1088/1126-6708/2008/04/034, arXiv:arXiv:0802.2879.
- [49] G. Polesello and D. R. Tovey, “Supersymmetric particle mass measurement with the boost-corrected contranverse mass”, *JHEP* **1003** (2010) 030, doi:10.1007/JHEP03(2010)030, arXiv:0910.0174.
- [50] O. Buchmueller, M. J. Dolan, and C. McCabe, “Beyond Effective Field Theory for Dark Matter Searches at the LHC”, *JHEP* **1401** (2014) 025, doi:10.1007/JHEP01(2014)025, arXiv:1308.6799.
- [51] O. Buchmueller, M. J. Dolan, S. A. Malik, and C. McCabe, “Characterising dark matter searches at colliders and direct detection experiments: Vector mediators”, *JHEP* **1501** (2015) 037, doi:10.1007/JHEP01(2015)037, arXiv:1407.8257.
- [52] A. Ballestrero et al., “PHANTOM: A Monte Carlo event generator for six parton final states at high energy colliders”, *Comput.Phys.Commun.* **180** (2009) 401, doi:10.1016/j.cpc.2008.10.005, arXiv:0801.3359.
- [53] J. Alwall et al., “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”, *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.
- [54] C. Degrande et al., “Effective Field Theory: A Modern Approach to Anomalous Couplings”, *Annals Phys.* **335** (2013) 21, doi:10.1016/j.aop.2013.04.016, arXiv:1205.4231.
- [55] CMS Collaboration, “Measurement of the B(s) to  $\mu^+ \mu^-$  branching fraction and search for B0 to  $\mu^+ \mu^-$  with the CMS Experiment”, *Phys. Rev. Lett.* **111** (2013) 101804, doi:10.1103/PhysRevLett.111.101804, arXiv:1307.5025.
- [56] C. Collaboration, “Technical proposal for the upgrade of the CMS detector through 2020”,.
- [57] C. Collaboration, “Technical Proposal for the Phase-II Upgrade of the Compact Muon Solenoid”,.
- [58] CMS Collaboration, “CMS Phase II Upgrade Scope Document”, Technical Report CERN-LHCC-2015-019. LHCC-G-165, CERN, Geneva, Sep, 2015.

- [59] T. T. Böhlen, F. Cerutti, M. P. W. Chin, A. Fassò, A. Ferrari, P. G. Ortega, A. Mairani, P. R. Sala, G. Smirnov and V. Vlachoudis, “The FLUKA Code: Developments and Challenges for High Energy and Medical Applications”, *Nuclear Data Sheets* **120** (2014) 211–214.
- [60] A. Ferrari, P. R. Sala, A. Fassò and J. Ranft, “FLUKA: a multi-particle transport code”, CERN-2005-10, INFN/TC\_05/11, SLAC-R-773 (2005).
- [61] CMS Collaboration V. Karimki, et al., “The CMS tracker system project: Technical Design Report”. Technical Design Report CMS. CERN, Geneva, 1997.
- [62] CMS Collaboration, “Addendum to the CMS tracker TDR”,.
- [63] CMS Tracker Sensor Collaboration, “Campaign to identify the future CMS tracker baseline”, *Nucl.Instrum.Meth.* **A658** (2011) 30, doi:10.1016/j.nima.2011.05.028.
- [64] CMS Tracker Collaboration, “Characterisation of silicon sensor materials and designs for the CMS Tracker Upgrade”, *PoS Vertex2012* (2013) 016.
- [65] G. L. G.-L. Casse, A. A. Affolder, P. Allport, and M. Wormald, “Measurements of charge collection efficiency with microstrip detectors made on various substrates after irradiations with neutrons and protons with different energies”, in *Proceedings of Science, PoS(VERTEX 2008)*, p. 036. 2008.
- [66] A. Chilingarov, “Temperature dependence of the current generated in Si bulk”, *JINST* **8** (2013) P10003, doi:10.1088/1748-0221/8/10/P10003.
- [67] CERN, LpGBT specification document, <https://espace.cern.ch/GBT-Project/LpGBT/Specifications/LpGbtSpecifications.pdf>.
- [68] I. Rubinskiy, “An EUDET/AIDA Pixel Beam Telescope for Detector Development”, *Physics Procedia* **37** (January, 2012) 923–931, doi:10.1016/j.phpro.2012.02.434.
- [69] T. Obermann et al., “Implementation of a configurable FE-I4 trigger plane for the AIDA telescope”, *Journal of Instrumentation* **9** (2014), no. 03, C03035.
- [70] G. Bianchi for the CMS Collaboration, “tkLayout: a design tool for innovative silicon tracking detectors”, *JINST* **9** (2014) C03054, doi:10.1088/1748-0221/9/03/C03054.
- [71] S. Mersi, D. Abbaneo, N. De Maio, and G. Hall, “CMS Tracker layout studies for HL-LHC”, *Physics Procedia* **37** (2012) 1070, doi:10.1016/j.phpro.2012.03.729.
- [72] C. D. Jones, M. Paterno, J. Kowalkowski, L. Sexton-Kennedy, and W. Tanenbaum, “The New CMS Event Data Model and Framework”, *Proceedings of CHEP 2006*, volume 1, editor S. Banerjee (India: Macmillan), 248.
- [73] CMS Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, *JINST* **9** (2014) P10009, doi:10.1088/1748-0221/9/10/P10009.
- [74] The PETTL Collaboration, “Enabling Technologies for Silicon Microstrip Tracking Detectors at the HL-LHC”, arXiv:1604.08583.
- [75] J. Olsen, T. Liu, Y. Okumura, “A Full Mesh ATCA-based General Purpose Data 3 Processing Board”, FERMILAB-CONF-13-527-CMS-PPD.

- [76] E. Flumerfert et al., “Off-the-Shelf DAQ”.
- [77] CMS Collaboration Collaboration, “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV”, Technical Report arXiv:1607.03663. CERN-PH-EP-2015-305. CMS-JME-13-004, CERN, Geneva, Jul, 2016. Comments: Submitted to J. Instrum. All the figures and tables can be found at <http://cms-results.web.cern.ch/cms-results/public-results/publications/JME-13-004/>.
- [78] D. Felici et al., “A 20 mW, 4.8 Gbit/sec, SEU robust serializer in 65nm for read-out of data from LHC experiments”, *Journal of Instrumentation* **9** (2014), no. 01, C01004.
- [79] T. Roy et al., “QIE: performance studies of the next generation charge integrator”, *Journal of Instrumentation* **10** (2015), no. 02, C02009.
- [80] CMS Collaboration, “Radiation Testing of Electronics for the CMS Endcap Muon System”, *Nucl. Instrum. Meth.* **A698** (2013) 242.
- [81] CMS Collaboration, “Results of radiation tests of the cathode front-end boards for the CMS endcap muon chambers”, *Nucl. Instrum. Meth.* **A471** (2001) 340.
- [82] B. Bylsma et al., “Results of radiation test of the cathode front end board for CMS endcap muon chambers”, in *Proceedings, 6th Workshop on Electronics for LHC experiments, Cracow, Poland, 11-15 Sep 2000*, pp. 231–235. 2000.
- [83] CMS Collaboration, “Aging tests of CMS muon chamber prototypes”, *Nucl. Instrum. Meth.* **A488** (2002) 240.
- [84] CMS Collaboration, “CMS The Muon Project Technical Design Report”. CERN-LHCC-1008-032, CMS-TDR-3, 1997.
- [85] B. G. Bylsma et al., “The cathode strip chamber data acquisition electronics for CMS”, *Nucl. Instrum. Meth.* **A600** (2009) 661–672, doi:10.1016/j.nima.2008.12.141.
- [86] CMS Collaboration, “The performance of the CMS muon detector in proton-proton collisions at  $\sqrt{s}=7$  TeV at the LHC”, *JINST* **8** (2013) P11002, doi:10.1088/1748-0221/8/18/P11002.
- [87] “XILINX Corp.:Soft Error Mitigation (SEM) Core”, Technical Report <http://www.xilinx.com/products/intellectual-property/sem.html>, 2014.